IDENTIFICATION OF BATCH GRINDING MILL - A COMPARATIVE STUDY

A Thesis submitted
In Partial Fulfilment of the Requirements
For the Degree of
MASTER OF TECHNOLOGY

by

R. SANKARAN

to the

Department of Electrical Engineering
Indian Institute of Technology Kanpur

AUGUST - 1972

DEDICATED TO

THE MEMORY OF

MY MOTHER

C E R T I F I C A T E

Certified that this work on "Identification of Batch Grinding Mill - A comparitive Study" by Mr. R. Sankaran has been carried out under my supervision and this has not been submitted elsewhere for a degree.

P. Ramakrishna Rao.

DR. P. RAMAKRISHNA RAO
ASSISTANT PROFESSOR
DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
KANPUR

A C K N O W L E D G E M E N T

The author wishes to express his indebtedness and sincere thanks to Dr. P.R.K. Rao for the continuous help, inspiration and guidance that he received from him during the course of this work. The author is thankful to Dr.P.C. Kapur for proposing an interesting and practical problem and for the helpful discussions he had with him. The facilities provided in the IBM 7044 system in the Computer Centre are gratefully acknowledged. Thanks are due to Mr. Yogendra Chandra for his patient typing and members of the Department of Electrical Engineering for their cooperation in preparing the thesis.

TABLE OF CONTENTS

A B S T R A C T

Various procedures for the off-line identification of a Batch Grainding Mill in noisy environment are discussed. A computer simulation of the mill with the identified parameters reveals the suitability of these schemes in a practical situation. A comparitive evaluation is then carried out and conclusions drawn.

CHAPTER I

INTRODUCTION

1.1 Outline of the Problem

Modelling and identification of parameters of the system so modelled is normally the first step in the theoretical investigation of the performance of a physical system and is of cruicial importance in any optimization study or adaptive control of the same. Identification calls for input-state or input-output records of the relevant system and therefore involves experimental work followed by theoretical analysis.

This thesis is concerned with a specific problem in Metallurgical Engineering which is deemed to belong to the area of identification. The system considered pertains to the model of a batch grinding ball mill which receives dolomite stones of specified size range and as the mill revolves, these are ground to finer sizes by the crushing and grinding action of steel balls which are mixed with the feed. The grinding action is assumed to be described by an apriori known set of linear differential equations as discussed subsequently. Experimental measurement of the distribution of particle sizes for different revolutions of the mill represents the data. Although dynamics of the plant and the measurement process can only be satisfactorily

characterised statistically, as is usual in many a physical situation, no statistical information concerning these is available. Since the structure or topology of the model and the nature of variations of parameters are assumed, the problem reduces to one of identifying these parameters to satisfy a chosen performance criterion.

This problem is characterised by the following features:

1. The dynamics of the mill are characterised by a continuous model, but measurements are discrete.

2. Strictly speaking, 2 sets of parameters, the so called "selection functions" and the "breakage functions" are to be identified. Though the model is linear in state variables, the measurements are non linearly related to the parameters. This thereby leads to the nonlinear parameter observability problem.

3. The data available is meagre and is not regularly spaced along the time axis. No ensemble of data is available and so the accuracy of any one reading is quite limited. To be fair, there are many practical problems in collecting accurate data at closer intervals as this is the case where measurements disturb the process.

1.2 Outline of the Dissertation

Different schemes for the identification of the system parameters have been considered in the following

chapters. On the whole it has not been possible to give any analytical proof for the convergence or uniqueness of the procedures employed and numerical convergence has been considered satisfactory.

Chapter II considers the mathematical model of the grinding process in terms of the selection and breakage functions. The model is derived using phenomenological considerations and the selection and breakage parameters themselves are affected probably in a complicated way by many factors like the properties of the material ground, the size of steel balls, the proportion of the balls in relation to the feed mass and so on. It is not an easy task to take into account any of these in our problem. Essentially we judge the success of the modelling by the closeness with which the experimental data could be simulated analytically.

Chapter III covers the first method-Recursive least squares. Here the performance measure is the well known sum of squares of errors between the simulated model and the experimental results. It is seen that the method has not been successful in solving the problem completely. While it does give a reasonably satisfactory answer to the question of identifying the selection functions, the same is

not true as far as breakage functions are concerned. This incidentally brings to sharp focus the type of difficulties and the lack of any general approach to the nonlinear parameter observability problem. What is suitable for one problem may not be so good in a slightly different situation.

Chapter IV discusses the Quasilinearization technique as applied to the identification problem using the continuous model. The criterion is same as in Chapter III and the algorithm provides more or less consistent results for both sets of parameters. Finally a complete program for the evaluation of the best parameter of the system given the data is presented.

We discuss the Invariant Imbedding approach to the identification problem in Chapter V. The many practical difficulties encountered in implementing the estimation algorithm for a problem of a specific nature are pointed out. Computational results for a set of initial conditions are presented and the limitations analysed.

Chapter VI covers the classical Maximum Likelihood Identification scheme with Gaussian assumptions. Here a more elaborate model with both driving and measurement noise is identified and is compared with previous results.

In addition this method provides an estimate of the various noise and error covariances for the assumed model.

A simple Stochastic Approximation algorithm is implemented in chapter VII which is essentially the Keifer Wolfowitz scheme with a suitable performance criterion The formulation is analogous to the steepest descent technique in the deterministic situation.

Chapter VIII concludes the dissertation with a comparitive evaluation of the different methods considered earlier.

CHAPTER II

THE MATHEMATICAL MODEL OF THE GRINDING MILL

Generally speaking, the mathematical modelling of the process of crushing and grinding of feed in a mill can be extremely involved depending upon the desired accuracy and details. Accordingly many models have been proposed by various authors (Schuhmann, 1940), (Charles 1957). In this work the phenomenological model derived by conceiving two basic functions in grinding is used. These are termed the selection and breakage functions and are defined subsequently. This approach has been found to be generally satisfactory for characterising and simulating of mills of the type considered. In this scheme, a set of time differential mass balance equations are presented (Reid; 1965)

$$\frac{dx_i(t)}{dt} = - S_i\, x_i(t) + \sum_{j=1}^{i-1} S_j\, Z_{ji}\, x_j(t),$$

$$i = 1, 2, \ldots\ldots n \qquad (2.1.1)$$

where $x_i(t)$ is the mass fraction of particles identified by the i th size index at time t and n is the number of particle size indices catagorized by two consecutive meshes of standard testing screens. S_i, the selection function is the rate of breakage of particles of i th size expressed in mass fraction ground per minute. Z_{ji}, the breakage

function is the i th size fraction of products when particles of size index j are broken.

In compact form eqn. 1 can be written as

$$\frac{dx(t)}{dt} = Ax(t) \qquad (2.1.2)$$

where x(t) is the 12 dimensional fractional particle size vector and A is 12 x 12 matrix given by

$$A = \begin{bmatrix} -S_1 & 0 & \cdots & 0 \\ S_1Z_{12} & -S_2 & 0 \cdots & 0 \\ S_1Z_{13} & S_2Z_{23} & -S_3 \cdots & 0 \\ \vdots & \vdots & & \vdots \\ S_1Z_{112} & S_2Z_{212} & \cdots & -S_{12} \end{bmatrix}$$

It may be noted that the above equation is a convenient approximation to a process which is possibly nonlinear and which at the moment does not admit of any dynamical noise. An additional error of course is introduced at the measurement stage. In this model altogether there are 78 unknown parameters - 12 selection functions and 66 breakage functions. It is felt that this large number of unknown parameters precludes any pratical solution of the identification problem without further simplifying assumptions.

There exists considerable experimental justification to assume that the breakage function is normalizable. (Herbst & Fuerstenau, 1968), (Brown 1969). This essentially means that for particles distributed according to Tylor mesh scale it can be assumed that $Z_{ji} = Z_{i-j}$, $i > j$. This assumption reduces the number of breakage parameters to just 11 and the total number of unknowns to 23. Of the two sets of parameters, the selection function S is relatively more important, both theoritically and practically.

The following points motivate a stochastic interpretation to the problem of identification as applied to the grinding mill.

1. The linear time invariant model discussed above is only a convenient approximation to the grinding process, and hence is open to question. Accordingly, the whole exercise may be viewed as one which fits the best model belonging to the class of linear time invariant systems into the experimental data and can be interpreted as a parameter optimization problem with an a priori chosen performance criterion.

2. It is not clear whether any additional disturbance entering the system dynamically will be able to "explain" the data better. The best that can be said here

is that it is one possible approach and consequently calls for a statistical formulation.

3. The major factor which justifies a stochastic interpretation readily is the presence of possible measurement errors. Also it is worthwhile to note the readings can be critically influenced by speed variations of the mill and the exact number of revolutions that has been completed. In a practical situation instant acceleration and abrupt stopping of the mill are not attainable.

It is quite possible that most of these errors are, to a first degree approximation, zero-mean, and hence can be averaged out by an ansemble of measurements by repeated experimentation. However, it is noted that such an ensemble is not available at the moment for reasons which are not entirely obscure.

The experimental data used in this work for identifying the parameters is provided by Berlioz (Berlioz L.M. 1966) and is considered to be reasonably accurate among the many data that are available. The particle size distribution for three feed masses given by Berlioz are tabulated in tables 2.1, 2.2 and 2.3.

TABLE 2.1

BERLIOZ DATA : MILL FEED 3300 GMS.

Size Index	Mesh Size	MILL REVOLUTIONS 0	20	40	60	80	100	150	200	300
1	8x10	1.0	.7797	.6216	.4964	.3888	.3103	.1764	.1155	.0419
2	10x14	0.0	.0933	.1343	.1622	.1725	.1706	.1426	.1225	.0512
3	14x20	0.0	.0472	.0828	.1076	.1223	.1329	.1399	.1271	.0883
4	20x28	0.0	.0226	.0429	.0610	.0822	.0965	.1186	.1243	.1071
5	28x35	0.0	.0147	.0280	.0412	.0566	.0680	.0942	.1061	.1206
6	35x48	0.0	.0097	.0192	.0304	.0398	.0487	.0781	.0893	.1106
7	48x65	0.0	.0079	.0160	.0242	.0317	.0403	.0526	.0694	.1094
8	65x100	0.0	.0044	.0113	.0173	.0221	.0298	.0445	.0522	.0749
9	100x150	0.0	.0045	.0069	.0114	.0175	.0206	.0310	.0380	.0598
10	150x200	0.0	.0030	.0064	.0093	.0133	.0154	.0242	.0335	.0505
11	200x270	0.0	.0025	.0050	.0075	.0097	.0127	.0187	.0239	.0352
12	270x400	0.0	.0023	.0044	.0069	.0093	.0114	.0171	.0213	.0333

TABLE 2.2

BERLIOZ DATA : MILL FEED 1980 GMS.

Size Index	MILL REVOLUTION 0	20	40	60	80	100	150	200	300
1	1.00	.7298	.5460	.4037	.2997	.2228	.0857	.0396	.0085
2	0.00	.1090	.1488	.1612	.1550	.1380	.0960	.0536	.0051
3	0.00	.0551	.0944	.1203	.1341	.1378	.1027	.0686	.0266
4	0.00	.0294	.0536	.0744	.0907	.1015	.1281	.1318	.0595
5	0.00	.0184	.0257	.0542	.0698	.0831	.1108	.1178	.0899
6	0.00	.0129	.0265	.0392	.0521	.0641	.0936	.1109	.1128
7	0.00	.0104	.0217	.0337	.0458	.0579	.0789	.1003	.1366
8	0.00	.0075	.0153	.0228	.0316	.0424	.0652	.0860	.1172
9	0.00	.0053	.0115	.0187	.0217	.0305	.0491	.0678	.0808
10	0.00	.0050	.0091	.0134	.0205	.0214	.0479	.0489	.0728
11	0.00	.0024	.0066	.0096	.0130	.0166	.0278	.0347	.0504
12	0.00	.0033	.0080	.0128	.0144	.0189	.0307	.0325	.0565

TABLE 2.3

BERLIOZ DATA : MILL FEED 3960 GMS.

Size Index	MILL REVOLUTIONS 0	20	40	60	80	100	150	200	300
1	1.0	.7960	.6718	.5229	.4655	.3806	.2351	.1438	.0651
2	0.0	.0882	.1089	.1615	.1727	.1796	.1625	.1287	.0776
3	0.0	.0407	.0737	.0982	.1092	.1249	.1399	.1339	.1049
4	0.0	.0238	.0435	.0612	.0712	.0865	.1134	.1289	.1265
5	0.0	.0133	.0261	.0382	.0439	.0568	.0812	.1020	.1208
6	0.0	.0086	.0174	.0385	.0295	.0416	.0589	.0777	.1018
7	0.0	.0068	.0134	.0206	.0220	.0227	.0480	.0651	.0847
8	0.0	.0046	.0091	.0153	.0180	.0225	.0338	.0449	.0653
9	0.0	.0035	.0069	.0108	.0140	.0175	.0251	.0365	.0625
10	0.0	.0030	.0068	.0085	.0105	.0140	.0212	.0220	.0416
11	0.0	.0020	.0035	.0051	.0091	.0099	.0149	.0196	.0327
12	0.0	.0021	.0039	.0066	.0071	.0101	.0130	.0135	.0263

CHAPTER III

RECURSIVE LEAST SQUARES

3.1 Introduction

In this Chapter the first method for identifying the selection and breakage functions so as to meet a chosen criterion is considered. This is the well known least square error measure which, historically is one of the earliest methods of estimating states and parameters. It is simple, mathematically tractable and conceptually quite appealing. The popularity rests partly on the essential fact that no statistics about the disturbances affecting the system be known.

The system description chosen is particularly simple. ie, no driving noise in the dynamics of the plant is assumed and the errors in data are deemed to be solely due to measurement noise.

3.2 Recursive Least Squares via Discrete Model

As seen in the last chapter the system is described by

$$\dot{x}(t) = A\,x(t)$$

and

$$y(t) = x(t) + v(t)$$

where $x(t)$ and $y(t)$ are 12 dimensional state and measurement vectors of particle sizes, $v(t)$ is the measurement

noise and A is 12 x 12 matrix given by

$$A = \begin{bmatrix} -S_1 & 0 & 0 & \cdots\cdots & 0 \\ S_1Z_1 & -S_2 & 0 & \cdots\cdots & 0 \\ S_1Z_2 & S_2Z_1 & -S_3 & \cdots\cdots & 0 \\ \vdots & & & & \vdots \\ S_1Z_{11} & S_2Z_{10} & S_3Z_9 & \cdots\cdots & -S_{12} \end{bmatrix}$$

Strictly speaking the identification of the plant matrix A consists of estimating both S and Z vectors simultaneously. This develops computationally into a difficult problem. Apart from the larger order of simultaneous equations considered, (23 eqns. corresponding to both S and Z vectors), the differing sensitivity of these two sets of parameters to the output data and the further accenuating of the nonlinearity of the parameters-observations relationships create difficulties in convergence. This is partly overcome by a step-by-step process of keeping the Z vector at a suitable constant value while iterating for S vector and vice-versa. For the success of this procedure it is essential that there should be monotonic convergence towards the least square solution.

3.3 Recursive Algorithm for S Vector

The algorithm for updating S given below assumes a constant Z for successive iterations. The essential idea

is to start with a nominal value of the parameters and develop an updating prodedure which recursively minimises the sum of squares of errors over the entire data in a sequential manner.

In the following no probabilistic description of the system initial condition is assumed, in accordance with the claim that in the Berlioz data of grinding, the feed consists of particle size of index one only. Any error implied by this assumption is considered negligible.

The continuous model of the plant in terms of the nominal S and Z vectors is given by

$$\dot{x}^\circ(t) = A^\circ x^\circ(t) \qquad (3.3.1)$$

$$y(t) = x^\circ(t)$$

$$\text{with } x^\circ(t_o) = (1, 0 \ldots\ldots 0)^T \qquad (3.3.2)$$

where $x^\circ(t)$, $y(t)$ and $x^\circ(t_o)$ are of dimension 12.

Taking the discrete equivalent, 1 and 2 give, for a discretizing time interval T, corresponding to 20 revolutions of the mill.

$$x^\circ(k+1) = e^{A^\circ T} x^\circ(k), \quad k=1, \ldots 5 \qquad (3.3.3)$$

$$x^\circ(k+1) = e^{2.5A^\circ T} x^\circ(k), \quad k=6,7 \qquad (3.3.4)$$

and

$$x^\circ(k+1) = e^{5.0A^\circ T} x^\circ(k), \quad k=8 \qquad (3.3.5)$$

$$y^{\circ}(k) = x^{\circ}(k),\ k=1,\ 2\ \ldots\ldots\ 9 \qquad (3.3.6)$$

Let there exist an admissible lower triangular matrix A which is stable such that the system given by

$$x(k+1) = e^{AT}x(k)$$
$$y(k) = x(k) + v(k) \qquad (3.3.7)$$

with initial condition $x(t_o) = x^{\circ}(t_o)$, where $y(k)$, $k = 1, 2 \ldots\ldots 9$ indeed is the given experimental data, such that the sum of squares of errors, $\sum_{k=1}^{9} v^T(k)\, v(k)$ is minimum with respect to the admissible class of A matrices.

It may be noted that in eqn. 6 and 7 the vectors $x(k)$, $x(k+1)$, $v(k)$ and the matrix A are all unknown. This beings out the crucial assumption to be made in a noisy identification problem where the parameters and hence the state trajectory are not known a priori and we are left with the noise corrupted measurements of inputs and outputs only. Accordingly we express A in the form $A = A^{\circ} + \Delta A$ with the implicit assumption that A° is "sufficiently close" to A so that $x^{\circ}(k)$ is a good approximation to $x(k)$ for all k. Analytically, how close should be A° to A is a ~~particularly~~ difficult question to answer. It evidently depends on the signal to noise ratio, and the sensitivity of the system

states to changes in parameter values. In a specific problem like this, numerical experimentation offers the best means of choosing A°. From eqns. 6 and 7,

$$y(k+1) = e^{(A^\circ+\Delta A)T} x^\circ(k) + v(k+1)$$

Expanding and retaining only the first 2 terms,

$$y(k+1) = e^{A^\circ T}\left[I+\Delta AT\right] x^\circ(k) + v(k+1),\ k=1,\ldots5 \quad (3.3.8)$$

Here again our assumption that ΔA is "small" comes into play. In this context, one particularly disturbing thought is that in our problem the time interval T has been fixed already by the availability of experimental data, making it beyond our control, and moreover T is not constant during the entire length of process. The eventual culmination of this is discussed later on.

It may also be noted that in the derivation k is limited to 5 only because of non-uniform step size beyond 100 revolutions of the mill. Necessary multiplicative factors are used in the computation later on to account for this.

Rewriting 8,

$$\begin{aligned} y(k+1) &= e^{A^\circ T}\left[I+\Delta AT\right] x^\circ(k) + v(k+1) \\ &= F^\circ \left[I+\Delta AT\right] x^\circ(k) + v(k+1) \qquad (3.3.9) \\ &\qquad k = 1,2, \ldots\ldots5 \end{aligned}$$

ie, $y(k+1) - F^{\circ}x^{\circ}(k) = F^{\circ}\Delta A T\, x^{\circ}(k) + v(k+1)$

where

$$\Delta A = \begin{bmatrix} -\Delta S_1 & 0 & 0 & \cdots & 0 \\ \Delta S_1 Z_1 & -\Delta S_2 & 0 & \cdots & 0 \\ \Delta S_1 Z_2 & \Delta S_2 Z_1 & -\Delta S_3 & \cdots & 0 \\ \vdots & & & & \vdots \\ \Delta S_1 Z_{11} & \Delta S_2 Z_{10} & \Delta S_3 Z_9 & \cdots & -\Delta S_{12} \end{bmatrix}$$

ie, $y(k+1) - F^{\circ}\, x^{\circ}(k) = H(k)\Delta S + v(k+1)$ (3.3.10)

where $H(k)\Delta S = F^{\circ}\Delta A T\, x^{\circ}(k)$, $k=1, \ldots 5$

Here H(k) may be interpreted as the modulation matrix (12 x 12) which relates in a linear way the correction ΔS to the L.H.S. which is a function of the measurements and the nominal trajectory only.

Letting $r(k) = y(k+1) - F^{\circ}x^{\circ}(k)$ and renaming $v(k+1)$ as $v(k)$,

$$r(k) = H(k)\Delta S + v(k), \quad k = 1, \ldots 8 \qquad (3.3.11)$$

which is the linearized equation sought all along in terms of ΔS.

Giving equal weightage to errors in all particle sizes the performance index to be minimised is

$$J = \sum_{k=2}^{9} v^{T}(k)\, v(k)$$

In the actual procedure this minimization of J is carried out sequentially as seen later on. However, for a single stage of the process, let

$$J(k) = v^T(k)\, v(k), \qquad k=2, \ldots 9$$

$$= \left[r(k)-H(k)\Delta S_k\right]^T\left[r(k)-H(k)\Delta S_k\right]$$

Minimising w.r.t. ΔS_k, the necessary condition to be satisfied is $$\frac{dJ(k)}{d\Delta S_k} = 0$$

which yields $$\Delta S_k = \left[H^T(k)H(k)\right]^{-1} H^T(k)r(k) \qquad (3.3.12)$$

provided the required inverse exists. This implies that H(k) be of full rank.

Now, considering a multistage process, let r(k+1) be the additional "observation" given by

$$r(k+1) = H(k+1)\Delta S + v(k+1) \qquad (3.3.13)$$

If we were to couple equations 11 and 13 with ΔS now denoted by ΔS_{k+1} to indicate that measurements upto the k+1st stage have been taken into consideration, we get

$$\Delta S_{k+1} = \left[\begin{Bmatrix} H(k) \\ H(k+1) \end{Bmatrix}^T \begin{Bmatrix} H(k) \\ H(k+1) \end{Bmatrix}\right]^{-1} \begin{bmatrix} H(k) \\ H(k+1) \end{bmatrix}^T \begin{bmatrix} r(k) \\ r(k+1) \end{bmatrix} \qquad (3.3.14)$$

It may be noted that the approach necessitates inversion of matrices of ever increasing order at every step of measurement and no use is made of our previous calculation ΔS_k. Computationally the implications are quite serious; we are solving the entire problem all over again, whenever a new set of measurements are available.

To avoid this situation and to process the data sequentially we use the Matrix Inversion Lemma (Sage 1968). The lemma is given below and is easily proved.

If $P_{k+1}^{-1} = P_k^{-1} + H^T(k+1)\, H(k+1)$, then

$$P_{k+1} = P_k - P_k\, H^T(k+1) \left[I + H(k+1) P_k H^T(k+1)\right]^{-1} H(k+1) P_k \qquad (3.3.15)$$

where all required inverses exist.

A few words about the notation used may be appropriate at this stage. When the index k is used as an argument, for example H(k), it indicates the variable at the k th stage only. On the other hand when used as a subscript, for example P_k or ΔS_k we imply the variable to be the result of calculations utilizing all the stages i, $1 \leq i \leq k$.

Rewriting eqns. 12 and 14,

$$\Delta S_k = P_k\, H^T(k)\, r(k) \qquad (3.3.16)$$

$$\text{and} \quad \Delta S_{k+1} = P_{k+1}\left[H^T(k)r(k)+H^T(k+1)r(k+1)\right] \qquad (3.3.17)$$

where $P_k^{-1} = H^T(k)H(k)$

and $P_{k+1}^{-1} = H^T(k)H(k)+H^T(k+1)H(k+1)$

Substituting for $H^T(k)r(k)$ in terms of P_{k+1} obtained by definition of the latter into eqn. 17 we get

$$\Delta S_{k+1} = P_{k+1}\left[P_{k+1}^{-1}-H^T(k+1)H(k+1)\right]\Delta S_k+P_{k+1}H^T(k+1)r(k+1)$$

$$\text{ie,}\ \Delta S_{k+1}=\Delta S_k+P_{k+1}H^T(k+1)\left[r(k+1)-H(k+1)\Delta S_k\right] \qquad (3.3.18)$$

Also by using the Matrix Inversion Lemma,

$$P_{k+1} = P_k-P_kH^T(k+1)\left[I+H(k+1)P_kH^T(k+1)\right]^{-1}H(k+1)P_k \qquad (3.3.19)$$

The algorithm is implemented with starting values for S and Z and ΔS calculated sequentially taking into account the entire data. The starting values are now updated by $S \leftarrow S+\Delta S$ and the entire procedure repeated. In every case the nominal solution using the discrete model and initial conditions was obtained. The sum of squares of errors involved is also calculated and in fact gives the clue to convergence or otherwise of the problem. The results are tabulated in table 3.1. It is seen that we get fairly satisfactory convergence of all selection functions for different starting values.

3.4 Results

The convergence of the selection functions S_1, S_2, S_{11} and S_{12} which were considered representative, are shown in fig. 3.1 and 3.2 for different starting values. The latter also shows the minimization of the sum of squares of error for a close starting value. (case I of table 3.1) It may be noted that even for this close initial guess, the convergence is slow and oscillatory.

The same approach is next tried for updating the value of Z, the breakage function. It is found that the procedure is not successful as there is no convergence even with close starting values. The behaviour is oscillatory with eventual divergence.

It is apparent that on the whole the data is not sufficient to draw any conclusion straightaway and such an attemptmay be presumptuous. But numerically the trouble is mainly about the lower elements of the Z vector which get modified disproportionally resulting in build up of errors. It appears that for the discrete model the sensitivity of the different components of the Z vector to the data is widely different. It may also be noted that the large and non-uniform discretization of the system can introduce appreciable error in the algorithm which neglects the higher order terms

in eqn. 3.3.8. This is indirectly verified in the quasi-linearization technique where we use the continuous model and sampling is done at closer intervals.

To summarise the positive and negative aspects of the above method, we have a workable algorithm for the identification of the selection functions when there exists a reasonably good knowledge of the breakage parameters. In this connection it is pointed out that the latter can be estimated very simply with due assumptions on the precision of the readings at 20 revolutions of the mill as follows:

By definition Z_{j-i} is the i th size fraction of products when particles of size index j are broken. Putting j = 1 and i = 2, 3 ..... 12, and considering a short interval of the mill with feed of only size index 1, we get

$$Z_i = \frac{x_{i+1}}{1-x_1} \qquad i = 1,2,3 \ldots\ldots 11,$$

both x_{i+1} and x_1 measured at 20 revolutions of the mill. The above relation generally suffers from the limited accuracy of a particular set of measurements and the assumption that secondary breakage of daughter particles can be neglected in the interval.

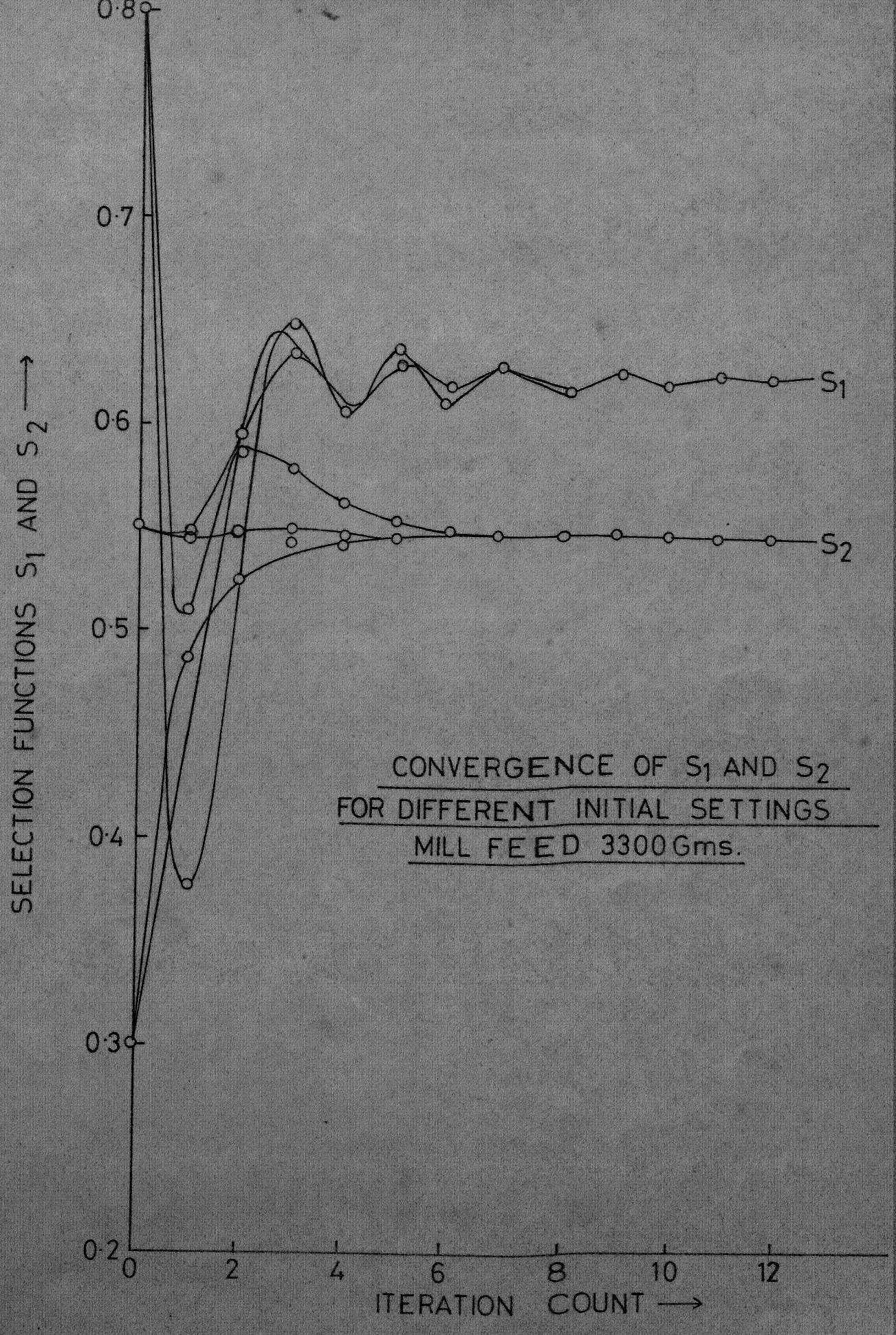
In fact the readings for the convergence of S shown in table 3.1 correspond to this particular value of Z. When compared with the refined parameter estimates given by the Quasilinearization method, what is obtained here by the above, it is seen, is only marginally inferior.

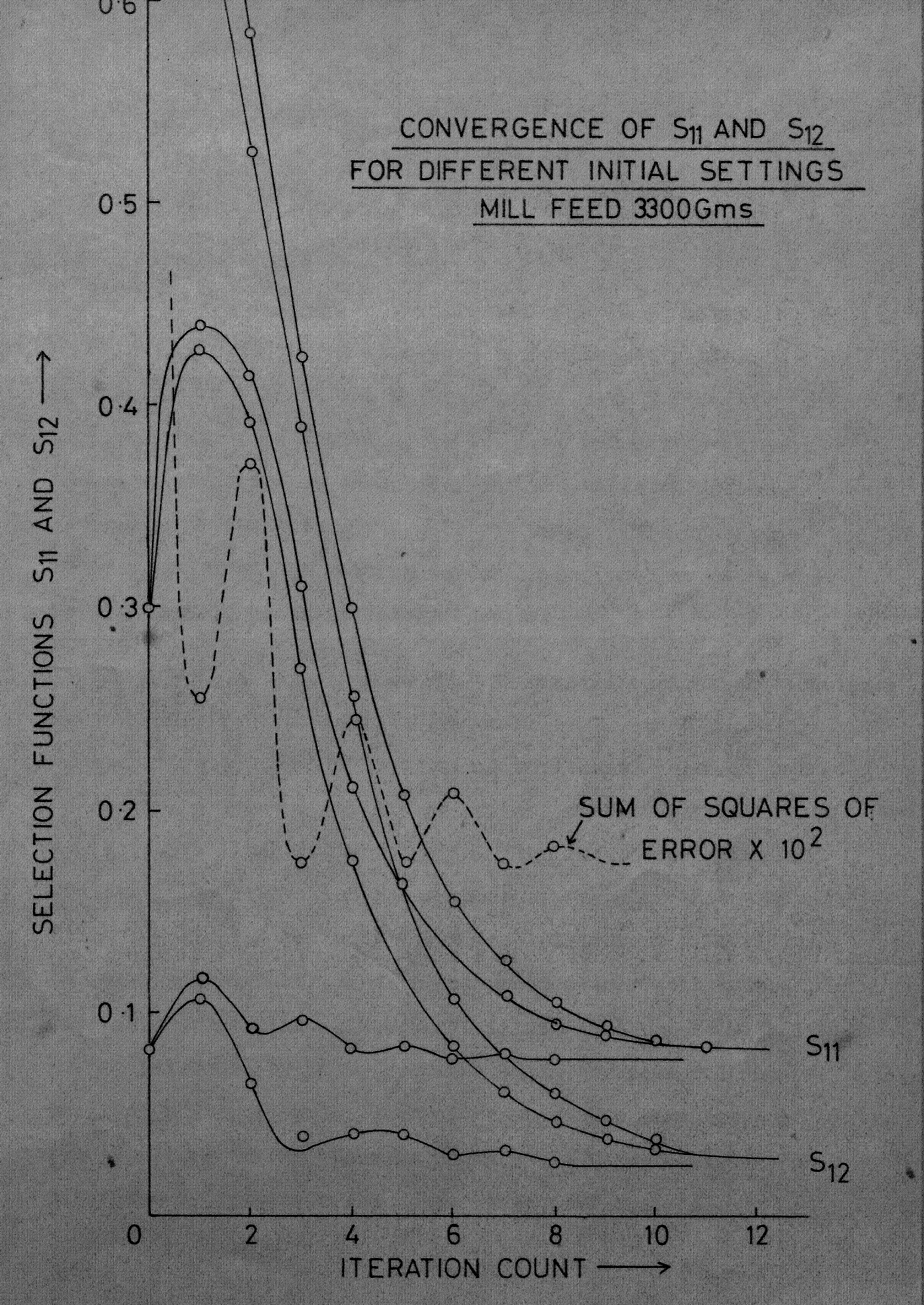
TABLE 3.1

CONVERGENCE OF SELECTION FUNCTIONS FOR DIFFERENT STARTING VALUES

Size Index	Breakage Function	Case I		Case II		Case III	
		Start	At End of 8 Itrns.	Start	At End of 10 Itrns.	Start	At End of 12 Itrns.
1	.42351	0.55	.6203	0.3	.6211	0.8	.6217
2	.21425	0.55	.5457	0.3	.5462	0.8	.5459
3	.10259	0.40	.4416	0.3	.4430	0.8	.4421
4	.06673	0.35	.3462	0.3	.3494	0.8	.3474
5	.04403	0.30	.2518	0.3	.2549	0.8	.2515
6	.03586	0.25	.1908	0.3	.1942	0.8	.1901
7	.01997	0.20	.1046	0.3	.1086	0.8	.1040
8	.02043	0.15	.1008	0.3	.1051	0.8	.1012
9	.01362	0.12	.0950	0.3	.0990	0.8	.0962
10	.01135	0.10	.0482	0.3	.0537	0.8	.0495
11	.01044	0.08	.0761	0.3	.0801	0.8	.0793
12	--	0.08	.0273	0.3	.0334	0.8	.0295
Sum of Squares of Errors x 10^4		180.60	18.53	4312.8	18.09	796.8	17.92

MILL FEED 3300 GMS.


0·8
0·7
0·6
0·5
0·4
0·3
0·2
SELECTION FUNCTIONS S_1 AND S_2 ⟶
S_1
S_2
0
2
4
6
8
10
12
ITERATION COUNT ⟶
CONVERGENCE OF S_1 AND S_2
FOR DIFFERENT INITIAL SETTINGS
MILL FEED 3300 Gms.


CONVERGENCE OF S_{11} AND S_{12}
FOR DIFFERENT INITIAL SETTINGS
MILL FEED 3300Gms
SELECTION FUNCTIONS S_{11} AND S_{12} ⟶
0·6
0·5
0·4
0·3
0·2
0·1
0
2
4
6
8
10
12
ITERATION COUNT ⟶
SUM OF SQUARES OF ERROR X 10^2
S_{11}
S_{12}

CHAPTER IV

QUASILINEARIZATION VIA THE CONTINOUS MODEL

4.1 Introduction

The parameter identification problem except in trivial cases being essentially a nonlinear estimation problem which even in the determiniatic cases defies straight forward analytical solution, it is quite natural to attempt linearization about a chosen nominal parameter value in an endeavour to successively update the same. This approach has been found to be quite powerful in our problem and gives more or less very satisfactory results.

In a broad framework, the quasilinearization technique is essentially a generalised Newton-Raphson method for functional equations. The original nonlinear system is solved by solving a sequence of linear systems which when convergence occurs finally gives the solution for the original problem. The motivation for such an approach is that a nonlinear multipoint boundary value problem is very difficult to solve analytically, and is not particularly well suited for computer solution. However this is not the case with linear multipoint boundary value problems where the powerful superposition principle holds good and the problem lends it self for effective computational methods. The Quasilinearization algorithm is characterised by two essential features -

quadratic convergence and monotonic convergence. Consistency of the results is numerically verified by giving widely different starting values for the parameters to be identified.

The application of the method of quasilinearization to an identification problem consists of augmenting the states by expressing the (time invariant) parameters in the form of a trivially dynamic system. The "state equation" is now clearly non linear which is expanded in Taylor series about the nominal parameter values, retaining only the linear terms. At this stage what we obtain is a time-varying linear system with unknown initial conditions. In fact the missing initial conditions are determined to satisfy an apriori criterion that the solution should satisfy. And in this case the performance index is again the sum of squares of the errors over the entire data. The problem is formulated using the continuous model keeping the discretization interval free to be chosen in the solution. In common with the method of least squares, we assume no driving disturbance and are satisfied incorporating measurement noise only.

4.2 Derivation of the Quasilinearization Algorithm

As mentioned in section 3-2, the identification of parameters is done in 2 stages which follow each other in an alternative fashion.

a. An iterative procedure of estimating S vector with Z maintained constant throughout at a nominal value to begin with, which is subsequently updated by step b.

b. An almost identical procedure as above, with the performance index minimised with respect to Z in a sequential manner, keeping S at the best value obtained by step a.

Algorithm for Step a

We have the system

$\dot{x}(t) = A(S)\, x(t)$ where the 12x12 matrix A is known in terms of the unknown parameter S, and $x(t_o)$ is known.

$$y(t) = x(t) + v(t)$$

where $y(t)$ is the measurement vector

and $v(t)$ measurement noise.

Consider the N + 1st iteration of the algorithm, the iteration index being indicated by the right superscript.

$$\dot{x}^{N+1}(t) = A(S^{N+1})\, x^{N+1}(t), N = 1, 2, \ldots \qquad (4.2.1)$$

Expanding the R.H.S. about the nominal value S^N and $x^N(t)$,

$$\dot{x}^{N+1}(t) = A(S^N)x^N(t) + \left[\frac{\partial}{\partial x}\left\{A(S)x(t)\right\}_{\substack{x=x^N \\ S=S^N}}\right]\left[x^{N+1}(t)-x^N(t)\right]$$

$$+ \left[\frac{\partial}{\partial S}\left\{A(S)x(t)\right\}_{\substack{x=x^N \\ S=S^N}}\right]\left[S^{N+1} - S^N\right] \qquad (4.2.2)$$

$$= A(S^N)x^N(t) + A(S^N)\left[x^{N+1}(t) - x^N(t)\right]$$

$$+ L\left[x^N(t),\, S^N\right]\left[S^{N+1} - S^N\right] \qquad (4.2.3)$$

where $$L\left[x^N(t), S^N\right] = \frac{\partial}{\partial S}\left[A(S)x(t)\right]_{\substack{x(t)=x^N(t) \\ S \;\; =S^N}} \qquad (4.2.4)$$

In lieu of the a priori assumption that S is time independent, $\dot{S}^{N+1}(t) = 0$ (4.2.5)

Augmenting the states we get

$$\begin{bmatrix} \dot{x}^{N+1}(t) \\ \dot{S}^{N+1}(t) \end{bmatrix} = \left[\begin{array}{c|c} A(S^N) & L(S^N,t) \\ \hline 0 & 0 \end{array}\right] \begin{bmatrix} x^{N+1}(t) \\ S^{N+1}(t) \end{bmatrix} - \begin{bmatrix} L(S^N,t)S^N \\ 0 \end{bmatrix}$$

(24x1) (24x24) (24x1) (24x1)

(4.2.6)

In principle, the R.H.S. is fully known-both the parameters and states correspond to the N th iteration; and equation 6 can be solved for any initial condition. Formally writing down the solution,

$$\begin{bmatrix} x^{N+1}(t) \\ - - - \\ S^{N+1}(t) \end{bmatrix} = \left[\begin{array}{c|c} \phi_{11}(t, t_o) & \phi_{12}(t, t_o) \\ - - - - & - - - - \\ \phi_{21}(t, t_o) & \phi_{22}(t, t_o) \end{array}\right] \begin{bmatrix} x^{N+1}(t_o) \\ - - - \\ S^{N+1}(t_o) \end{bmatrix} + \begin{bmatrix} p_1(t) \\ - - - \\ p_2(t) \end{bmatrix} \tag{4.2.7}$$

where $p_1(t)$ and $p_2(t)$ are particular solutions of 6 and $\Phi(t, t_o)$ which is shown in partitioned form is the state transition matrix.

As far as the solution is concerned, it is clear that $p_2(t) \equiv 0$ and since $S^{N+1}(t) \equiv S^{N+1}(t_o)$ for all time t, $\phi_{21}(t, t_o) \equiv 0$ and $\phi_{22}(t, t_o) \equiv I$. So writing down the non-trivial part of eqn. 7, we have

$$x^{N+1}(t) = \phi_{11}(t, t_o)x^{N+1}(t_o) + \phi_{12}(t,t_o)S^{N+1}(t_o)+p_1(t) \tag{4.2.8}$$

$\phi_{11}(t, t_o)$ and $\phi_{12}(t,t_o)$ are obtained from the overall state transition matrix $\Phi(t,t_o)$ determined by

$$\dot{\Phi}(t, t_o) = \left[\begin{array}{c|c} A(S^N) & L(t) \\ \hline 0 & 0 \end{array}\right] \Phi(t,t_o) \tag{4.2.9}$$

with $\Phi(t_o,t_o) = I$

And finally $p_1(t)$ is given by the following differential equation

$$\dot{p}_1(t) = A(S^N)p_1(t) - L(t, S^N)S^N \qquad (4.2.10)$$

with initial condition $p_1(t_o) = 0$

Now the solution takes the form

$$x^{N+1}(t) = \phi_{11}(t,t_o)x^{N+1}(t_o)+\phi_{12}(t,t_o)S^{N+1}(t_o)+p_1(t) \qquad (4.2.11)$$

At this stage, the essentials of what follows can be summed up to see where all this ultimately leads. It may be emphasised here that $S^{N+1}(t_o)$ is still unknown.

In principle $\phi_{11}(t,t_o)$ and $\phi_{12}(t,t_o)$ and $p_1(t)$ can be solved from the known initial conditions and $x^{N+1}(t)$ for any t can be obtained in terms of the missing initial conditions $S^{N+1}(t_o)$.

Let these equations be of form

$$\begin{aligned} x^{N+1}(t_1) &= H_1 S^{N+1}(t_o) + w(t_1) \\ &\vdots \qquad\qquad \vdots \\ x^{N+1}(t_8) &= H_8 S^{N+1}(t_o) + w(t_8) \end{aligned} \qquad (4.2.12)$$

where $t_1, t_2 \ldots\ldots t_8$ correspond to the time instants at

which measurements are available and H_1, H_2 ..... H_8 are completely known. Hence assuming measurement errors, the problem reduces to the determination of $S^{N+1}(t_o)$ so that $x^{N+1}(t_1)$, ... , $x^{N+1}(t_8)$ fit into the given data in the least squares fashion.

It can be easily seen that the forgoing exercise will be computationally very tedious and time consuming. To briefly summarise these requirements we have to solve two sets of 12x12 simultaneous (one of them time varying) linear first order differential equations for the state transition matrix, and another set of 12 linear time-varying first order differential equations for the particular solution $p_1(t)$. This will enable one to determine H_i, $i = 1, 2 \ldots\ldots 8$ to further proceed with the least square evaluation of S^{N+1}.

It was not considered desirable to proceed as above, but instead exploit the condition that $S^{N+1}(t)$ is time invariant and can be related to $S^N(t)$ by a simple constant "error term" ΔS, which is consequently evaluated using essentially all the forgoing ideas. The following accordingly eliminates the necessity of calculating a time varying state transition matrix, but employs a numerical

integration procedure which is conceptually and computationally straight forward.

we have $\dot{x}^{N+1} = A(S^{N+1})\, x^{N+1}(t)$

$$= A(S^N)x^N(t) + \left[\frac{\partial}{\partial x}\left\{A(S)x(t)\right\}_{\substack{x=x^N \\ S=S^N}}\right]\left[x^{N+1}(t)-x^N(t)\right]$$

$$+\left[\frac{\partial}{\partial S}\left\{A(S)x(t)\right\}_{\substack{x=x^N \\ S=S^N}}\right]\left[S^{N+1} - S^N\right]$$

$$= A(S^N)x^N(t) + A(S^N)\left[x^{N+1}(t) - x^N(t)\right]$$

$$+ L(x^N(t), S^N)\ \Delta S^N \qquad (4.2.13)$$

where $L(x^N(t), S^N)$ is given by eqn. (4.2.4)

ie, $\dot{x}^{N+1}(t) = A(S^N)x^{N+1}(t) + L(x^N, S^N, t)\Delta S^N \qquad (4.2.14)$

with $x^{N+1}(t_o) = (1, 0 \ldots\ldots\ldots\ldots 0)^T$.

Let $\Phi(t,t_o)$ be the state transition matrix for the time invariant system 14 given by

$$\Phi(t,t_o) = \Phi(t-t_o) = e^{A(S^N)(t-t_o)} \qquad (4.2.15)$$

with $\Phi(t_o,t_o) = I$

$$x^{N+1}(t) = \Phi(t-t_o)x^{N+1}(t_o) + \int_{t_o}^{t} \Phi(t-\tau)L(\tau)\Delta S^N d\tau \quad (4.2.16)$$

At $t = t_1$ where $L(\tau) = L(x^N, s^N, \tau)$

$$x^{N+1}(t_1) = \Phi(t_1-t_o)x^{N+1}(t_o) + \int_{t_o}^{t_1} \Phi(t_1-\tau)L(\tau)\Delta S^N d\tau$$

ie, $$x^{N+1}(t_1) - \Phi(t_1-t_o)x^{N+1}(t_o) = H_1 \Delta S^N \quad (4.2.17)$$

Considering the observations $y(t) = x^{N+1}(t)+v(t)$ we rewrite 17,

$$y(t_1)-v(t_1) - \Phi(t_1-t_o)x^{N+1}(t_o) = H_1 \Delta S^N$$

Minimising $\| v(t_1) \|^2_{R^{-1}}$ where R is the covariance of the observation error gives,

$$\Delta S_1^N = \left[H_1^T R^{-1} H_1\right]^{-1} H_1^T R^{-1} \left[y(t_1)-\Phi(t_1-t_o)x^{N+1}(t_o) \right] \quad (4.2.18)$$

Considering subsequent observations at t_k, k=2, ...θ, we can obtain ΔS_k sequentially by the use of the matrix inversion lemma as in section 3.3. The sequential estimation equations are:

$$P_{k+1} = P_k - P_k H^T(k+1) \left[R(k+1)+H(k+1)P_k H^T(k+1)\right]^{-1} \quad (4.2.19)$$

$$H(k+1)P_k$$

$$\Delta S_{k+1} = \Delta S_k + P_{k+1} H^T(k+1) R^{-1}(k+1) \left[y(t_{k+1}) - \Phi(t_{k+1}-t_o) x^{N+1}(t_o) - H(k+1) \Delta S_k \right] \quad (4.2.20)$$

Algorithm for Step b.

Here the S parameters are kept constant and updating of Z parameters is done. The essential ideas remain as in the previous algorithm.

$$\dot{x}^{N+1}(t) = A(Z^{N+1})\, x^{N+1}(t)$$

$$= A(Z^N)x^N(t) + \left[\frac{\partial}{\partial x} \left\{ A(Z)x(t) \right\}_{\substack{x=x^N \\ Z=Z^N}} \right] \left[x^{N+1}(t) - x^N(t) \right]$$

$$+ \left[\frac{\partial}{\partial Z} \left\{ A(Z)x(t) \right\}_{\substack{x=x^N \\ Z=Z^N}} \right] \left[Z^{N+1} - Z^N \right]$$

ie, $\dot{x}^{N+1}(t) = A(Z^N)x^{N+1}(t) + M \Delta Z \quad (4.2.21)$

where $M(Z^N, x^N(t)) = \left[\frac{\partial}{\partial z} A(Z)x(t) \right]_{\substack{x=x^N \\ Z=Z^N}}$

The procedure is exactly similar to what was done previously. The H(i) matrices i = 1, ... 8 are of dimension 11x11.

4.3 Computation Considerations

Each iteration consists of the following steps.

1. Nominal solution of the system with assumed or updated values of the parameters. The discretization interval was chosen quite small (corresponding to 2.5 revolutions of the mill) to evaluate the integral in eqn. 4.2.16 numerically.

2. Determination of H(i), i = 1, 2 ..... 8 and sequential estimation of $\triangle S_k$ (or $\triangle Z_k$ as the case may be)using equns. 4.2.19 and 4.2.20 (or corresponding eqns.).

3. Updating the starting parameter values and return to step 1.

As was indicated earlier, the entire problem was tackled by alternating steps a and b. For this approach to be successful, it is necessary that we get monotonic convergence in both the steps a and b. This is first of all verified for different starting values for both S and Z parameters to be true. Consistent values for the parameters and the sum of error squares were obtained for widely different initial guesses. These are tabulated in tables 4.1 and 4.2. The behaviour of S_1, S_2, S_{11}, S_{12}, Z_1, Z_6 and Z_{11} are shown in figs. 4.1, 4.2 and 4.3.

The same program was used for alternate iteration of S and Z vectors with suitable control statements and hence a complete pack for the identification of all parameters in the least square fashion was developed. By a study of the rate of convergence of the algorithm it was decided to have 4 iterations per step for both parts a and b, which alternate each other. This was considered satisfactory to attempt overall rapid convergence. It was also arranged such that the program automatically switches itself from one step to the other if the "gain" in the performance index is smaller than a prescribed figure at the end of every iteration. The stopping criterion is a gain of less than 1% in the performance index for both S and Z parameters, or 16 iterations, which takes almost 8 minutes in Fortran. If necessary the program can be rerun with the last updated values of parameters as inputs. But this will be unlikely if the initial guesses are not off by more than 20%.

It is also desirable from the point of view of lesser no. of iterations that the initial starting value for Z is calculated by the approximate procedure given in chapter III. The results for the different feed masses are given in tables 4.3, 4.4 and 4.5.

Considering the accuracy with which the data has been fitted, we get the sum of squares of error for 3300 gms feed to be .001428 whereas the sum of squares of the data itself is 2.16401, representing an average error of 2.6%. This seems to be a considerable improvement over the values which were known previously. (Average error 5.9%). The procedure was tried for feed masses 1980 gms. and 3960 gms. with resulting average error 2.9% and 3.4% respectively. The monotonic convergence is shown for all the cases in fig.4.4.

4.4 Discussion

We considered the quasilinearization method for solving a multiple point boundary value problem in the least square manner. There are many points in this chapter in common with chapter III. The criterion is the same, and we essentially try to force the nominal solution furnished by the assumed parameters to approach the data in a least square manner by a recursive procedure. Even the first method implies a sort of linearization by expanding e^{AT} and retaining only the linear terms. The essential difference between the two methods is the nature of the system eqns. tackled. In the first we formulate the linearized system in discrete form, where the time

interval was beyond our choice. And mostly due to this, the first algorithm is not successful for the iteration of Z vector. This flaw is removed in the Quasilinearization algorithm which has rapid and monotonic convergence for both S and Z parameters. As far as the results are concerned it is noted that quasilinearization is decidedly superior in as much as convergence is monotonic. We get consistent results which compare favourably with that of chapter III.

TABLE 4.1

CONVERGENCE OF SELECTION FUNCTIONS FOR DIFFERENT INITIAL SETTINGS-MILL FEED 3300 GMS.

Size Index	Z	Case I		Case II		Case III	
		Start	At End of 8 Itrns.	Start	At End of 13 Itrns.	Start	At End of 12 Itrns.
1	.4235	.6095	.6395	0.80	.6403	0.30	.6391
2	.2143	.5453	.5452	0.80	.5409	0.30	.5490
3	.1026	.4675	.4432	0.80	.4433	0.30	.4400
4	.0607	.3828	.3413	0.80	.3419	0.30	.3404
5	.0440	.2983	.2548	0.80	.2549	0.30	.2550
6	.0359	.2189	.1877	0.80	.1876	0.30	.1879
7	.0199	.1504	.1160	0.80	.1157	0.30	.1160
8	.0204	.0970	.1002	0.80	.0999	0.30	.1000
9	.0136	.0633	.1082	0.80	.1078	0.30	.1083
10	.0113	.0539	.0588	0.80	.0583	0.30	.0589
11	.0104	.0741	.0918	0.80	.0911	0.30	.0920
12	--	.1294	.0440	0.80	.0432	0.30	.0441
Sume of Squares of Error x10^4		78.49	34.9	680.94	34.30	4615.50	36.0

TABLE 4.2

CONVERGENCE OF BREAKAGE FUNCTIONS FOR DIFFERENT INITIAL SETTINGS-MILL FEED 3300 GMS.

Size Index	S	CASE I Start	CASE I At End Of 14 Itrns.	CASE II Start	CASE II At End Of 12 Itrns.
1	.6095	0.25	.4386	.500	.4388
2	.5453	0.25	.2254	.300	.2250
3	.4675	0.25	.1075	.200	.1077
4	.3828	0.25	.0620	.100	.0621
5	.2983	0.25	.0397	.080	.0397
6	.2189	0.25	.0314	.070	.0314
7	.1504	0.25	.0177	.050	.0177
8	.0970	0.25	.0142	.040	.0142
9	.0633	0.25	.0159	.030	.0159
10	.0539	0.25	.0116	.020	.0116
11	.0741	0.25	.0150	.020	.0151
12	.1294	--	--	--	--
Sum of Squares of Error $\times 10^4$		3.2×10^4	64.89	2696.0	64.89

SELECTION AND BREAKAGE FUNCTION DETERMINED ITERATIVELY-3300 GMS.

Size Index	STARTING VALUES S	Z	S AT The End of 4 Itrns	Z AT The End of 8 Itrns.	S AT the End of 12 Itrns.	Z AT The End of 16 Itrns.
1	0.60	.4235	.6192	.4347	.6214	.4386
2	0.55	.2142	.5462	.2130	.5526	.2137
3	0.45	.1025	.4421	.1005	.4525	.1008
4	0.35	.0667	.3384	.0627	.3519	.0607
5	0.25	.0440	.2510	.0445	.2585	.0442
6	0.20	.0358	.1841	.0329	.1941	.0310
7	0.10	.0199	.1121	.0232	.1117	.0248
8	0.10	.0204	.0952	.0184	.1140	.0173
9	0.10	.0136	.1019	.0133	.1087	.0131
10	0.08	.0113	.0515	.0113	.0629	.0112
11	0.08	.0104	.0818	.0103	.0973	.0102
12	0.05	--	.0334	--	.0502	--
ERSUM x_{10}^4	38.43		19.25	16.23	15.05	14.10

ERSUM = Sum of Squares of Error ; Sum of Squares of Data = 2.1640

SELECTION AND BREAKAGE FUNCTION DETERMINED ITERATIVELY-3960 GMS.

Size Index	Starting Values S	Z	S At End of 4 th Itrn.	Z At End of 8 th Itrn.	S At End of 12 th Itrn	Z At End of 16 th Itrn.
1	0.60	.4323	.5372	.4278	.5304	.4325
2	0.55	.1995	.4465	.2030	.4538	.2061
3	0.45	.1166	.3529	.1150	.3638	.1122
4	0.35	.0652	.2635	.0637	.2657	.0611
5	0.25	.0421	.2054	.0446	.2006	.0451
6	0.20	.0335	.1620	.0314	.1629	.0302
7	0.10	.0225	.1221	.0230	.1134	.0232
8	0.10	.0171	.1062	.0172	.0997	.0172
9	0.10	.0147	.0745	.0140	.0666	.0136
10	0.08	.0098	.0821	.0094	.0645	.0093
11	0.08	.0102	.0565	.0101	.0311	.0101
12	0.06	--	.0748	--	.0398	--
ERSUM $\times 10^4$	165.6		36.92	35.02	34.91	34.48

ERSUM = Sum of Squares of Error ; Sum of Square of Data = 3.2128

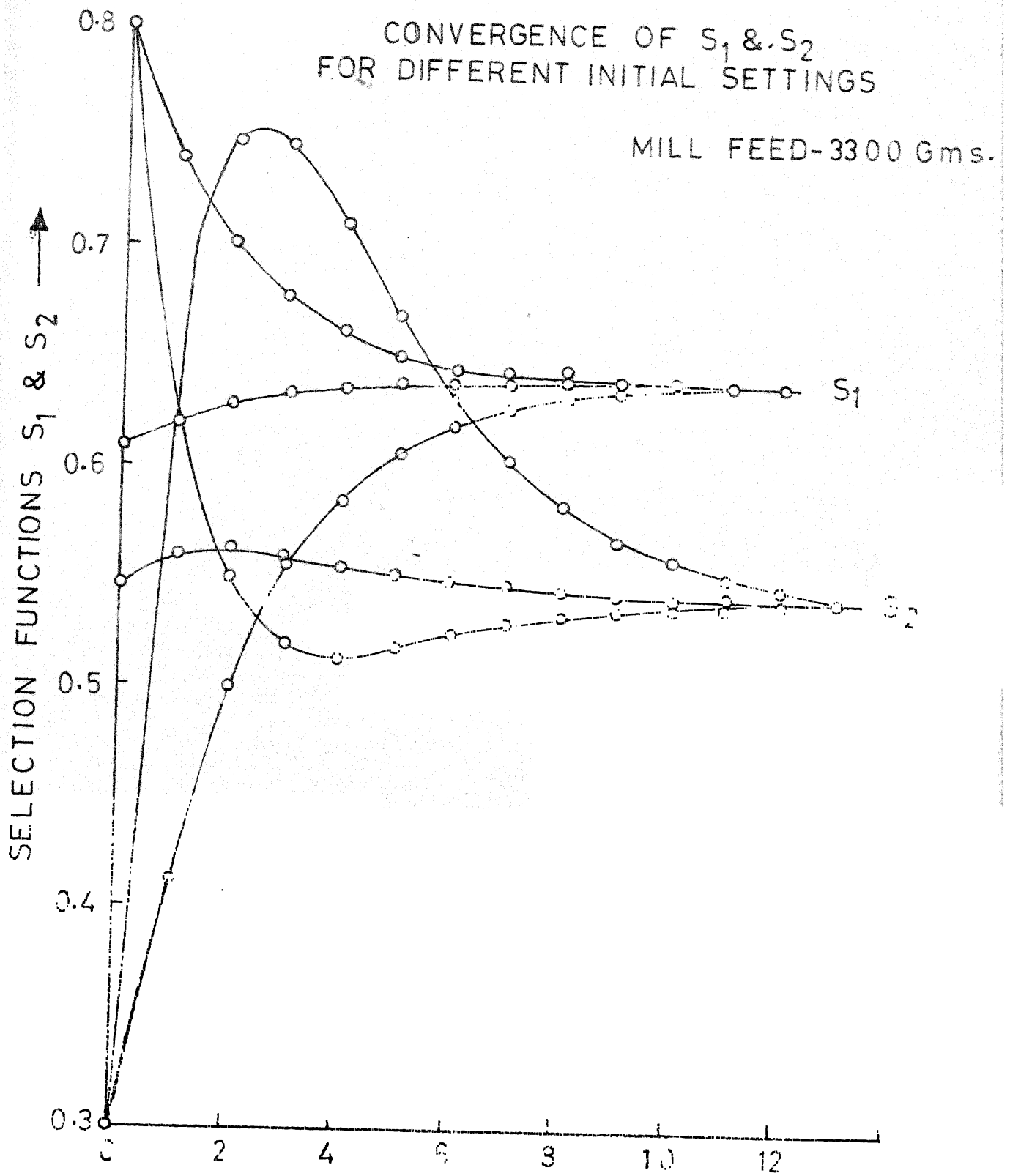

CONVERGENCE OF S_1 & S_2
FOR DIFFERENT INITIAL SETTINGS
MILL FEED-3300 Gms.
SELECTION FUNCTIONS S_1 & S_2
0·8
0.7
0.6
0.5
0.4
0.3
S_1
S_2
2
4
6
8
10
12
ITERATION COUNT

FIG. 4.1

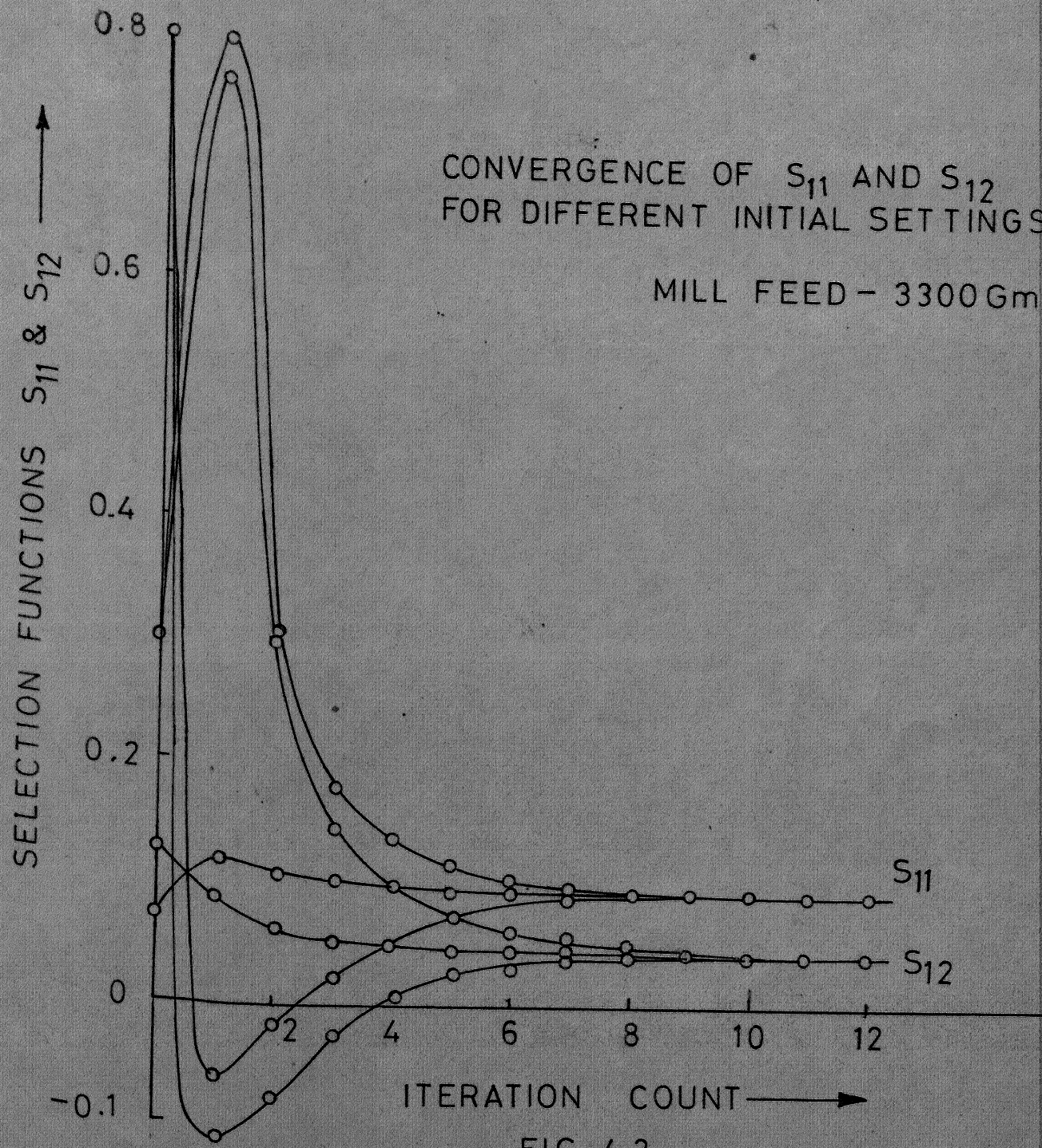

CONVERGENCE OF S_{11} AND S_{12}
FOR DIFFERENT INITIAL SETTINGS
MILL FEED - 3300 Gm
SELECTION FUNCTIONS S_{11} & S_{12}
0.8
0.6
0.4
0.2
0
-0.1
2
4
6
8
10
12
ITERATION COUNT
S_{11}
S_{12}

FIG. 4.2

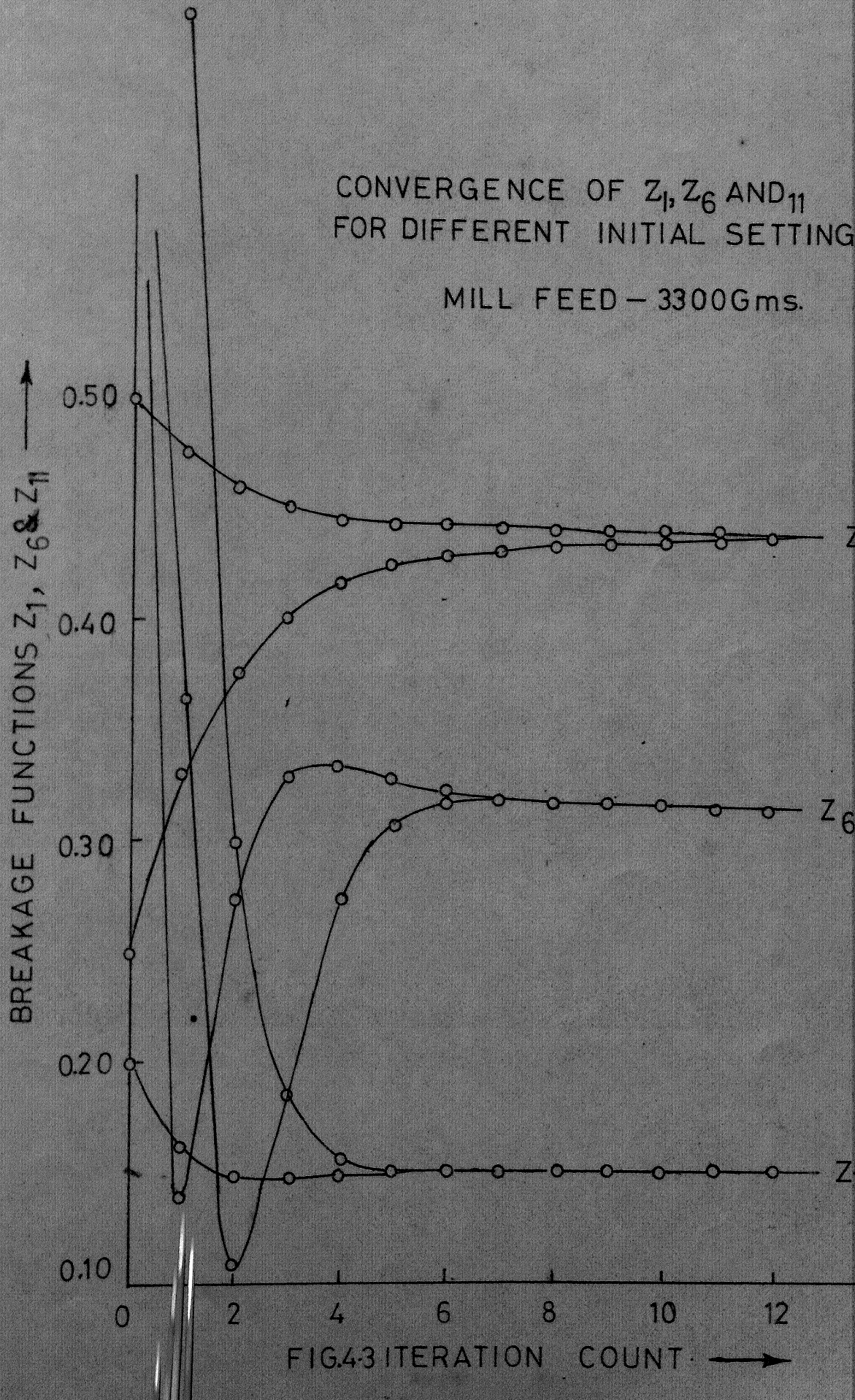

CONVERGENCE OF Z_1, Z_6 AND Z_{11}
FOR DIFFERENT INITIAL SETTING
MILL FEED – 3300Gms.
BREAKAGE FUNCTIONS Z_1, Z_6 & Z_{11}
0.50
0.40
0.30
0.20
0.10
0
2
4
6
8
10
12
ITERATION COUNT
Z_6

FIG.4.3

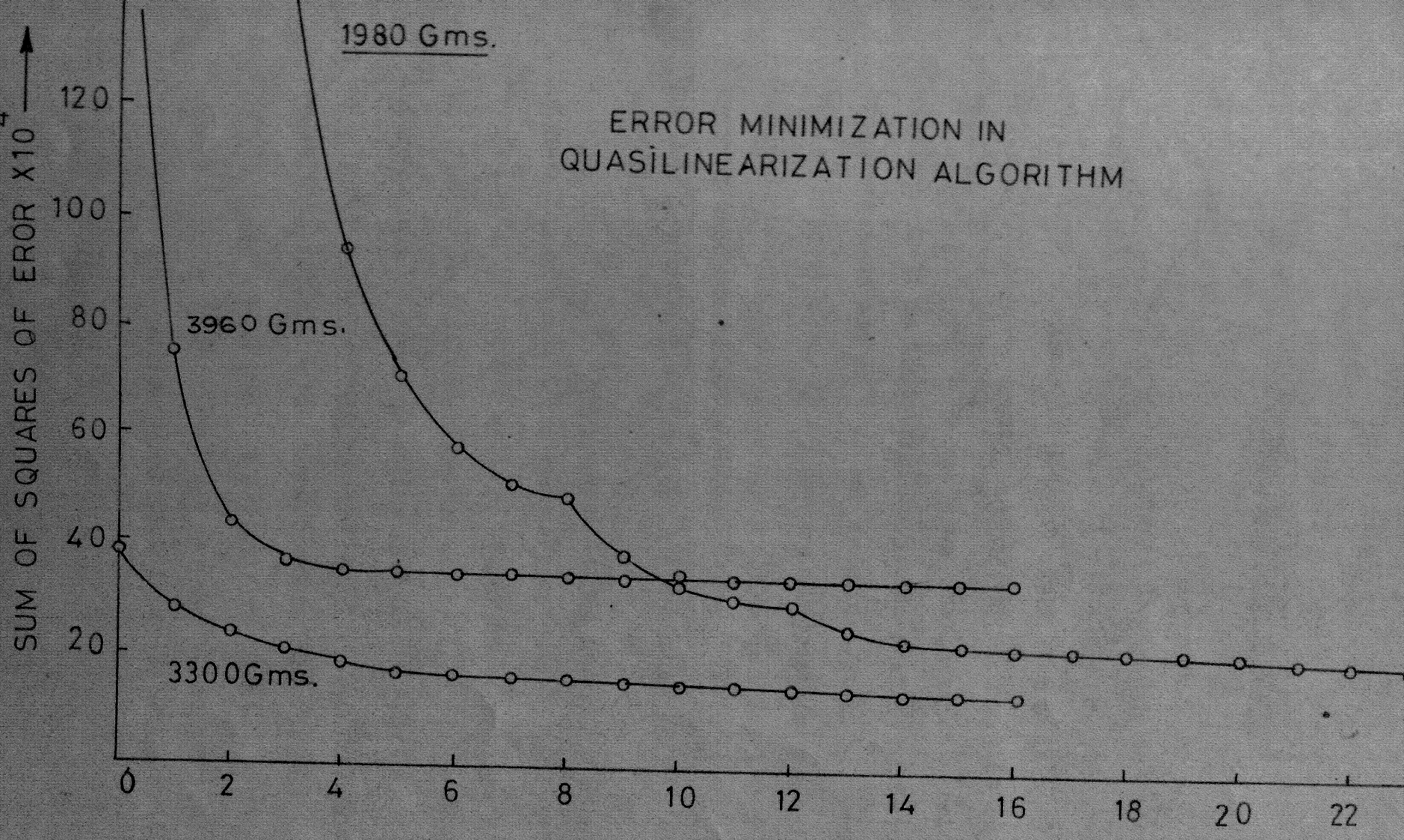

1980 Gms.
ERROR MINIMIZATION IN
QUASILINEARIZATION ALGORITHM
SUM OF SQUARES OF EROR X10^4
120
100
80
60
40
20
3960 Gms.
3300Gms.
0
2
4
6
8
10
12
14
16
18
20
22
ITERATION COUNT

FIG. 4.4

CHAPTER - V

METHOD OF CONTINUOUS INVARIANT IMBEDDING

5.1 Introduction

The mathematical model describing the entire process of grinding chosen in the previous chapters was particularly simple - it did not admit of any driving noise affecting the system in a dynamical way. It is true that so long as we do not have any knowledge of noise statistics, it is immaterial to put them at the driving end or measurement end of a linear dynamic system. Nevertheless it is interesting to conceive of a disturbance input for the system and reformulate the problem in a different way. This is done utilizing the Invariant Imbedding approach for the identification problem as discussed below.

In optimization studies, we generally encounter a two-point or multi-point boundary value problem where the terminal conditions are split. It may be emphasised here that the concept of invariant imbedding is quite general and powerful. Whereas quasilinearization represents an iterative procedure where a sequence of linear problems are solved, invariant imbedding tackles the original problem by expanding it into a family of problems. For example instead of considering a process of duration T the invariant imbedding formulation is to consider a set

of problems with the duration of the process ranging from zero to T. These problems are then imbedded to obtain the particular original problem.

The invariant imbedding algorithm for identification is formulated in terms of the continuous grinding model of chapter II. As can be seen later, in the algorithm that we obtain, by virtue of the moderately large dimension of both the system variables and unknown system parameters, the solution is very cumbersome in terms of the number of simultaneous non-linear differential equations to be solved. Hence in order to limit the computational requirements to reasonable levels it is assumed that the breakage functions are known within close range and identification of selection functions only is called for.

One more aspect of the problem, ie, the non-availability of readings along the time axis at uniform intervals is of great relevance here. In order that the algorithm may be implemented with reasonable accuracy, it is imperative to have measurements at closer intervals than is available in Berlioz data. Hence it becomes almost unavoidable to resort some from of interpolation of the given readings.

5.2 The Invariant Imbedding Algorithm

As in the quasilinearization technique we augment the states of the dynamic system by expressing the constant selection parameters as trivially dynamic. In the following derivation the vector x represents this augmented state of the system. Hence we now have

$$\dot{x}(t) = g(t, x) + k(t, x)\, u\ (t)$$
$$y(t) = h(t, x) + v(t)\ , \quad 0 \leq t \leq T \qquad (5.2.1)$$

where x is a n vector

y is m vector of observations

v is m vector of measurement error

u is unknown noise assumed to be a r vector

g(t,x) is a known n vector function

h(t,x) is known m vector function

k(t,x) is known nxr matrix function

To obtain a logical performance index we define the vector valued residual errors as

$$e_1(t) = y(t) - h(t, \hat{x}) \qquad (5.2.2)$$
$$e_2(t) = \dot{\hat{x}}(t) - g(t, \hat{x}) \qquad (5.2.3)$$

where $\hat{x}(t)$ is the estimate of x(t), $0 \leq t \leq T$.

Using the least square philosophy, the performance index is now

$$J = \int_0^T \left\| e_1(t) \right\|_Q^2 + \left\| e_2(t) \right\|_W^2 \, dt \qquad (5.2.4)$$

which is to be minimised with respect to $\hat{x}(t)$, $0 \leq t \leq T$. $Q(t, \hat{x})$ and $W(t, \hat{x})$ are at least positive semidefinite weighting matrices.

Substituting equations 1,2,3 into eqn. 4,

$$J = \int_0^T \left\| y(t) - h(t, \hat{x}) \right\|_Q^2 + \left\| \hat{u}(t) \right\|_{k^T W k}^2 \, dt \qquad (5.2.5)$$

where the minimisation is now with respect to both $\hat{x}(t)$ and $\hat{u}(t)$, $0 \leq t \leq T$ subject to the differential constraint

$$\dot{\hat{x}} = g(t, \hat{x}) + k(t, \hat{x})\, \hat{u}(t) \qquad (5.2.6)$$

Let $v(t, \hat{x}) = k^T(t, \hat{x})\, W(t, \hat{x})\, k(t, \hat{x})$

Define the Hamiltonian $H(t, \hat{x}, p, \hat{u})$

by $$H(t,\hat{x},p,\hat{u}) = \left\| y(t)-h(t,\hat{x}) \right\|_Q^2 + \left\| \hat{u} \right\|_V^2 + \left\langle p, g(t,\hat{x})+k(t,\hat{x})\hat{u} \right\rangle \qquad (5.2.7)$$

where p is the n dimensional co-state variable.

The unknown input $\hat{u}$ can be eliminated from eqn. 5.2.7 by the Pontryagin's Minimum Principle

$$\frac{\partial H}{\partial \hat{u}} = 0$$

which should be satisfied along the optimal trajectory which is now denoted x^* and the corresponding Hamiltonian as H^*.

Carrying out the elimination we get

$$H^*(t,x^*,p^*) = \left\| y(t)-h(t,x^*) \right\|_Q^2 + \langle p^*,\ g(t,x^*) \rangle - \tfrac{1}{4} \langle p^*, kV^{-1}k^T p \rangle \qquad (5.2.8)$$

Now the reduced state costate equations that the optimal trajectory should satisfy the following canonical equations:

$$\dot{x}^* = \frac{\partial H^*}{\partial p^*} (t,\ x^*,\ p^*) \qquad (5.2.9)$$

$$\dot{p}^* = \frac{-\partial H^*}{\partial x^*} (t,x^*,p^*) \qquad (5.2.10)$$

Since T has been fixed an $x^*(0)$ and $x^*(T)$ are free, the transversality condition yields

$$p^*(0) = 0\ ,\ p^*(T) = 0 \qquad (5.2.11)$$

At this stage the sequential estimation problem reduces to one of solving the two point boundary value problem characterised by eqns. 9 and 10 with boundary conditions of 11. It may be noted that both $x^*(0)$ and $x^*(T)$ are unknown. The invariant imbedding approach is now to consider a larger class of TPBVP's with boundary conditions.

$$p^*(0) = 0 \text{ and } p^*(T) = c \qquad (5.2.12)$$

where T is recognized as the running time variable and c a n vector. Now the missing terminal condition on x is denoted by r(c,T) where r is unknown. It can be easily be shown by perturbation techniques that r(c,T) satisfies the invariant imbedding equation (Detchmendy 1965, Sage 1968)

$$\frac{\partial r}{\partial T} - \frac{\partial r}{\partial c}\frac{\partial H^*}{\partial r}(T, r, c) = \frac{\partial H^*}{\partial c}(T, r, c) \qquad (5.2.13)$$

where $\left[\frac{\partial r}{\partial c}\right]_{ij} = \frac{\partial r_i}{\partial c_j}$

The above nonlinear partial differential equation is approximately solved by assuming a solution of form

$$r(t, T) = P(T)c + \hat{x}(T) \qquad (5.2.14)$$

where P(T) is an n x n matrix

Substituting eqn. 14 into eqn. 13,

$$\frac{dP}{dT}\, c + \frac{\partial \hat{x}}{\partial T} - P(T)\, \frac{\partial H^*}{\partial r}\,(T,\ Pc+\hat{x}, c) = \frac{\partial H^*}{\partial c}\,(T, Pc+\hat{x}, c) \qquad (5.2.15)$$

Next, we expand eqn. 15 about r(0,T) retaining terms to first order. The justification for this approach is that only those solutions of eqn. 13 for which c=0 are of interest to us. And the least square estimate of x(T) denoted by $\hat{x}(T)$ is r(0,T). After expanding eqn. 15, we collect terms of order c^0 (ie, independent of c), and c. The desired sequential estimator equations can be found by equating coefficient of c^0 and c to zero because c is completely arbitrary. After considerable algebra we get (Detchmandy 1965)

$$\dot{\hat{x}}(T) = g(\hat{x},T) + 2P(T)\left[\frac{\partial h^T(\hat{x},T)}{\partial \hat{x}}\right] Q \left[y(T)-h(\hat{x},T)\right] \qquad (5.2.16)$$

$$\dot{P}(T) = \frac{\partial g(\hat{x},\ T)}{\partial \hat{x}}\, P(T) + P(T)\, \frac{\partial g^T(\hat{x},\ T)}{\partial \hat{x}}$$

$$+ 2P(T) \left\{\frac{\partial}{\partial \hat{x}}\left[\frac{\partial h^T(\hat{x},T)}{\partial \hat{x}}\, Q \left\{y(T)-h(\hat{x},T)\right\}\right]\right\} P(T)$$

$$+ \tfrac{1}{2}\, k(T,\ \hat{x})\ V^{-1}(T,\ \hat{x})\ k^T(T,\ \hat{x}) \qquad (5.2.17)$$

Equations 16 and 17 represent the algorithm for the sequential estimation of the states and parameters of the system, taking into account measurements one by one. The major difficulty inherent in this approach is that the initial conditions $\hat{x}(0)$ and $P(0)$ are not known and is subsequently discussed in the next section.

5.3 Computational Considerations

Since the continuous imbedding formulation of the problem is utilised, we require data that is more closely spaced than is available in Berlioz's. A linear interpolation of the data is affected to obtain readings at every 5 revolutions of the mill.

Coming to the implementation of equns. 16 and 17, what we have is a pair of coupled multidimensional nonlinear ordinary first order differential equations. For a n dimensional system with m unknown parameters, a total of $(n + m) + (n + m)^2$ coupled nonlinear first order differential equations have to be solved with unknown initial conditions. Hence computationally this method is not well suited for a problem of even moderate dimension. Convergence is not assured for an arbitrary starting guess and the system of equns. 16 and 17 become unstable for

large values of the weighting matrices. Consequently whatever knowledge we already have about the system were used in choosing the initial conditions. It was also decided to give weightage to measurement errors so that any trend in the behaviour of the parameter estimates in approaching the results of the quasilinearization algorithm can be used as a cross check.

1. The initial condition of the states was taken to be the initial measurement itself, in common with previous methods.

2. Starting values for S were chosen about 25% - to 30% higher than the known values, as given by chapter IV.

3. Taking into account the fact that the matrix P is analogous to the error covariance of the states, the initial condition P(0) was chosen to be the variance of the states and parameters.

4. The elements of the weighting matrix Q (chosen to be diagonal) are determined by systematic trial and error, the idea being that for a chosen initial conditions the parameter estimates should tend to the known values as rapidly as possible and the estimate of the states should track the measurements more or less closely. It is found

that a Q matrix with the following diagonal elements in the given order satisfies the requirements reasonably well : 20.0, 200.0, 160.0, 160.0, 100.0, 100.0, 40.0, 40.0, 40.0, 35.0, 25.0, 25.0.

5.4 Conclusions

The integration is performed using second order predictor-corrector method. The results are shown in tables 5.1 and 5.2. Fig.5.1 gives the sequential estimation of the parameters S_1, S_2, S_3, S_{11} and S_{12}. It is seen by comparison with table 2.1 that the algorithm tracks the measurements closely except near the start. To check the reasonableness of the final set of selection functions furnished by the method, the program was rerun feeding back the final set of parameters as initial setting. (with corresponding modification of initial variances). The weighting matrix was unaltered. This final set of parameters is shown in the last column marked RUN II in table 5.2. And the estimate for the parameters nearly converges to that given by chapter IV. Though the result seems to be satisfactory, it has to be conceded that we had a good knowledge of the parameters from the very beginning which decidedly influenced the initial condition settings and weighting factors.

To summarise, it is doubtful whether the invariant imbedding approach will be successful in a similar problem where data is limited, no noise statistics are available and knowledge of the parameter values is poor. Basically, lack of knowledge about the initial condition and weighting factors for the coupled non-linear differential equations to be solved is a distinctive disadvantage for numerical solution of the same. Also numerical experimentation with different initial conditions is unattractive for a higher dimensional problem. Finally, in the case of a system with no driving noise, the basic idea of introducing one, and then optimizing an objective function using the Minimum Principle does not seem justifiable.

TABLE 5.1

ESTIMATION OF STATES BY INVARIANT IMBEDDING-MILL FEED 3300 GMS.

Size Index	MILL REVOLUTIONS: 0	20	40	60	80	100	150	200	300
1	1.0	.7639	.6118	.4932	.3903	.3099	.1788	.1143	.0429
2	0.0	.0936	.1378	.1620	.1724	.1713	.1430	.1215	.0534
3	0.0	.0503	.0831	.1074	.1227	.1326	.1389	.1274	.0882
4	0.0	.0257	.0444	.0616	.0808	.0959	.1179	.1242	.1070
5	0.0	.0164	.0290	.0416	.0561	.0681	.0938	.1063	.1199
6	0.0	.0114	.0200	.0303	.0399	.0489	.0771	.0896	.1102
7	0.0	.0086	.0162	.0242	.0318	.0402	.0532	.0694	.1087
8	0.0	.0059	.0113	.0173	.0223	.0294	.0444	.0525	.0748
9	0.0	.0042	.0076	.0114	.0171	.0208	.0309	.0382	.0595
10	0.0	.0035	.0065	.0094	.0132	.0157	.0242	.0334	.0504
11	0.0	.0029	.0051	.0076	.0098	.0126	.0187	.0247	.0352
12	0.0	.0026	.0046	.0068	.0093	.0114	.0170	.0214	.0332

TABLE 5.2

IDENTIFICATION OF SELECTION FUNCTIONS - MILL FEED 3300 GMS.

Size Index	Start	20	40	60	80	100	150	200	300	RUN II
1	0.80	.7652	.7032	.6683	.6595	.6531	.6373	.6258	.6179	.6016
2	0.70	.6572	.5871	.5503	.5444	.5439	.5523	.5241	.5483	.5542
3	0.60	.5656	.5055	.4719	.4665	.4627	.4526	.4434	.4457	.4469
4	0.50	.4657	.4053	.3721	.3636	.3575	.3462	.3338	.3436	.3443
5	0.35	.3284	.2905	.2694	.2645	.2613	.2541	.2466	.2460	.2412
6	0.30	.2786	.2404	.2194	.2145	.2113	.2035	.1959	.1957	.1905
7	0.15	.1425	.1292	.1218	.1200	.1189	.1163	.1136	.1136	.1118
8	0.15	.1424	.1291	.1218	.1201	.1190	.1164	.1137	.1134	.1114
9	0.15	.1423	.1291	.1217	.1201	.1188	.1164	.1137	.1135	.1115
10	0.10	.0935	.0821	.0758	.0744	.0734	.0712	.0689	.0689	.0670
11	0.10	.0956	.0881	.0839	.0829	.0822	.0808	.0792	.0791	.0780
12	0.08	.0746	.0651	.0598	.0586	.0579	.0560	.0541	.0539	.0525

Final value of Performance Index = 0.15346

CHAPTER VI

MAXIMUM LIKELIHOOD IDENTIFICATION

6.1 Introduction

One of the most widely accepted methods of estimation of non-random parameters in a dynamical system under noisy environment is the maximum likelihood estimation scheme. Indeed, it can be interpreted even in the case of random parameters as a form of Baye's estimator where no apriori knowledge of the parameter statistics is available. The basic idea is that any reasonable estimate of a parameter is that value which causes the given observation most likely, ie, that parameter value which maximises the conditional probability density induced on the observations. This estimate is called the maximum likelihood estimate.

Briefly the approach consists of obtaining the joint conditional density of the observations given the paramete θ in accordance with the dynamical model and noise structure chosen.

This gives the likelihood function which is subsequently maximised subject to the system equations. When successful, the method yields second order statistics of the disturbances also.

Although in the following derivation it is assumed that the noise distribution is Gaussian, it is,

rather inappropriate in the sense that in our problem the probability of obtaining any negative reading is zero and hence the spread of the disturbance has to be restricted. But mathematical convenience dictates the choice of a Normal probability density function for the noise sequences.

6.2 Reformulation and Solution of the Problem

The approach below treats the identification problem using the discrete model of the plant and associated disturbances. Here we use the matrix A as the discrete state transition matrix for a small but finite duration T of the process. As in the previous chapters, the process is characterised by a set of linear stationary first order simultaneous differential equations. It is also assumed that the noise sequences are uncorrelated Gaussian random processes.

Accordingly the system equations are given by

$$x(k) = Ax(k-1) + w(k-1)$$
$$y(k) = x(k) + v(k) \quad k = 2, \ldots\ldots N \qquad (6.2.1)$$

where A is the unknown plant matrix to be identified, x(k) is the state and y(k) the measurement vectors (both of dimension r), v(k) and w(k) are stationary zero mean uncorrelated noise sequences of dimension r.

The second order statistics are given by

$$E\left\{ w(i)\, w^T(j)\right\} = F\delta_{ij} \qquad (6.2.2)$$

and

$$E\left\{ v(i)\, v^T(j)\right\} = R\delta_{ij} \qquad (6.2.3)$$

where δ_{ij} is the Kronecker Delta,

$$\delta_{ij} = 1 \quad \text{when} \quad i = j$$
$$\delta_{ij} = 0 \quad \text{when} \quad i \neq j$$

Also the driving and measurement noise sequences are mutually uncorrelated.

$$\text{ie,} \quad E\left\{ v(i)\, w^T(j)\right\} = 0 \quad \text{for all } i, j$$
$$\text{and} \quad E\left\{ v(i)\, y^T(j)\right\} = 0 \quad \text{for all } i, j$$

The Likelihood Function

We wish to determine the maximum likelihood estimate of the matrix A and the noise variances F and R. We denote the elements of A and the elements of B (a r x r matrix, which is introduced later) by the vector θ. The conditional probability density function $p\left[y(k)/y(j),\ 1 \leq j \leq k-1\ ;\ \theta\right]$ of the measurement at the k th stage given all previous measurements and parameters is Gaussian by the assumption on the noise statistics,

and is given by $N\left[\hat{y}(k/\theta),\ P(k)\right]$ where

$$\hat{y}(k/\theta) = E\left\{y(k)/y(j)\quad 1 \leq j \leq k-1\ ;\quad \theta\right\} \qquad (6.2.4)$$

$$\text{and } P(k) = E\left[\left\{y(k) - \hat{y}(k/\theta)\right\}\left\{y(k)-\hat{y}(k/\theta)\right\}^{T}\right] \qquad (6.2.5)$$

It may be noted that as more and more observations are processed, $\hat{y}(k)$ tends to the true value and the error convariance matrix $P(k)$ tends to a constant matrix P which is positive definite.

From the system equations 1 and 2, we can obtain a recursive relation for $y(k/\theta)$ as follows.(Kashyep, 1970)

$$\begin{aligned} x(k) &= A\left[y(k-1) - v(k-1)\right] + w(k-1) \\ y(k) &= A\ y(k-1) - Av(k-1) + w(k-1)+v(k) \\ &= Ay(k-1) + w_1(k) \end{aligned} \qquad (6.2.6)$$

where $w_1(k)$ consists of two noise sequences w and v.

We define a new noise sequence $w_2(k)$ given by

$$w_2(k) = e(k) - Be(k-1) \qquad (6.2.7)$$

where e(k) is the error involved in the estimate of y(k) and is a r dimensional vector

$$\text{ie,} \quad e(k) = y(k) - \hat{y}(k/\theta)$$

where $E\{e(k)\} = 0$ and $E\{e(k)\ e^T(j)\} = P\delta_{kj}$ (6.2.8)

Then replacing $w_1(k)$ by $w_2(k)$ in eqn. 6,

$$y(k) - Ay(k-1) = e(k) - Be(k-1) \qquad (6.2.9)$$

The noise sequences $w_1(k)$ and $w_2(k)$ have to satisfy equality of second moments

ie $E\{w_1(k)\ w_1^T(k)\} = E\{w_2(k)\ w_2^T(k)\}$

ie $ARA^T + F + R = P + BPB^T$

Therefore, $F = P - R + BPB^T - ARA^T$ (6.2.10)

Also $E\{w_1(k)w_1^T(k-1)\} = E\{w_2(k)w_2^T(k-1)\}$

ie, $AR = BP$; Therefore, $R = A^{-1}BP$ (6.2.11)

We shall subsequently estimate the "best" value of P and use the same for obtaining R and F.

The likelihood function for large N can be now written as

$$L_N(\theta, P) \triangleq \ln p\left[y(1), \ldots\ldots y(N)/\theta\right] \qquad (6.2.12)$$

$$= \ln p\left[y(1/\theta)\right] + \sum_{j=2}^{N} \ln p\left[y(j)/y(k)\ 1 \le k \le N-1\ ;\ \theta\right]$$

$$= -rN/2 \ln 2\pi - N/2 \ln \|\det P\| - \tfrac{1}{2} \sum_{j=1}^{N} \|y(j) - \hat{y}(j/\theta)\|_P^{-1}$$

$$= -\frac{rN}{2}\ln 2\pi - \frac{N}{2}\ln \| \det P \| - \tfrac{1}{2}\sum_{j=1}^{N} \| e(j/\theta) \|_{P^{-1}}$$

(6.2.13)

where $p[y(1/\theta)]$ is chosen arbitrarily.

At this stage the problem reduces to one of maximising $L_N(\theta, P)$ with respect to θ where $e(k/\theta)$ obeys eqn.9.

Estimation of P, A and B

Choosing nominal starting values for A and B matrices we can determine the error sequence e(k), k = 2, ... N. Consistant with the assumption that there is negligible error in the measurement of initial conditions we set e(1) = 0 to initiate the sequence e(k).

Maximising the likelihood function $L_N(\theta,P)$ w.r.t. P we have

$$\left.\frac{\partial L_N(\theta,P)}{\partial P}\right|_{P=\hat{P}(N)} = 0$$

Simplifying we get

$\hat{P}(N) \triangleq$ Best estimate of P using N measurements

$$= \frac{1}{N} \sum_{j=1}^{N} e(j/\theta)\, e^{T}(j/\theta) \qquad (6.2.14)$$

Now substituting back in the likelihood function we get

$$L_N(\theta, \hat{P}(N)) = -\frac{rN}{2}\left[1+\ln 2\pi\right] - \frac{N}{2} \ln \det\left[\hat{P}(N)\right] \qquad (6.2.15)$$

Now maximising R.H.S. of eqn.15 is equivalent to minimising det $\hat{P}(N)$ which is redefined as $J_N(\theta)$

Hence $$J_N(\theta) = \det\left[\hat{P}(N)\right]$$

$$= \det \sum_{j=1}^{N} e(j/\theta)\, e^{T}(j/\theta) \qquad (6.2.16)$$

Now the maximum likelihood estimates of A and B are obtained by minimising $J_N(\theta)$ w.r.t. A and B where $e(k/\theta)$ obeys the eqn.

$$e(k/\theta) - B\, e(k-1/\theta) = y(k) - A\, y(k-1) \quad k=2, \ldots N$$

In addition we have the physical constraint on A which must be lower traiangular.

ie, $$a_{ij} = 0 \quad j > i, \quad i = 1, 2 \ldots\ldots\ldots r \qquad (6.2.17)$$

The procedure now entails evaluation of the constrained partial derivatives $\partial J_N(\theta)/\partial A$ and $\partial J_N(\theta)/\partial B$ and is discussed subsequently.

At this stage a practical difficulty in implementing the scheme as above has to be considered. As can be seen from eqn. 16, the matrix $\hat{P}(N)$ will be singular for any $N < r$. One necessary (but not sufficient) condition for $\hat{P}(N)$ not to be ill conditioned is that N is very large compared to r. This is not satisfied in our problem and serious numerical errors crop up in the evaluation of $\det \hat{P}(N)$ and $[\hat{P}(N)]^{-1}$. Both have to be evaluated repeatedly in the iterative procedure. After considerable numerical experimentation using double precision, it was concluded that for a system of moderate dimension as this, the solution, if at all possible will be of little value because of computational errors. Accordingly the objective function to be optimized was modified to

$$J(\theta) = \sum_{j=2}^{N} e^T(j/\theta)\, e(j/\theta) \qquad (6.2.18)$$

which is again minimised subject to eqns. 9 and 17.

It is not claimed that minimising the sum of squares of error is the same analytically as minimising the determinant of error conviance matrix. But under the specific conditions imposed by the problem it seems best to go in for the alternative, since, intuitively the modified criterion seems to be reasonable.

The augmented function Ja with $\lambda(k)$ and μ as Legrange Multipliers is formed.

$$J_a(\theta) = \sum_{k=2}^{N} e^T(k)e(k) + \sum_{k=2}^{N} \lambda^T(k) \left[e(k) - B\, e(k-1) - y(k) + A\, y(k-1) \right] + \mu^T \underline{a} \qquad (6.2.19)$$

where $\underline{a}$ is a vector consisting of all super diagonal elements of A which are constrained to be zero.

The $\lambda(k)$ sequence and μ are evaluated by

setting $$\frac{\partial J_a}{\partial e(k)} = 0 \qquad (6.2.20)$$

and $$\frac{\partial J_a}{\partial \underline{a}} = 0 \qquad (6.2.21)$$

The first equation gives

$$\lambda(k) = -2\, e(k) + B^T \lambda(k+1) \qquad (6.2.22)$$

Also $\lambda(N+1) = 0$

Hence by integrating backwards $\lambda(k)$, $k = 2, \ldots\ldots N$ are evaluated.

The gradients $\partial J_a / \partial A$ and $\partial J_a / \partial B$ are evaluated by partial differentiation taking into account eqn 21 also alongwith.

Therefore, $$\frac{\partial J_a}{\partial A} = \sum_{k=2}^{N} \lambda(k)\, y^T(k-1)$$

with $$\left[\frac{\partial J_a}{\partial A}\right]_{ij} = 0 \qquad j > i \qquad i = 1, \ldots\ldots r \qquad (6.2.23)$$

and $$\frac{\partial J_a}{\partial B} = \sum_{k=2}^{N} -\lambda(k)\, e^T(k-1) \qquad (6.2.24)$$

This completes the derivation of the essential equations for the optimization procedure.

6.3 Computational Aspects

Because of the very nature of the formulation and solution as discussed above, it is apparent that meaningful results can be expected only when large data is available.

Moreover when measurements are available at non uniform intervals the computational algorithm becomes very complicated, with no closed form expressions available for the various gradients. In that case it becomes necessary to evaluate them numerically which takes a prohibitively long time. So in common with the invariant imbedding approach, it was decided to interpolate the data corresponding to [illegible] every 5 revolutions of the mill. Different interpolating schemes were considered and were compared with a simulated solution using the results of chapter IV. It was observed that linear interpolation involves minimum error followed by a quadratic polynomial, and hence the former was adopted.

The minimisation was done iteratively, starting with a nominal value for A and B. One iteration consists of evaluation of the objective function, the Lagrange multipliers $\lambda(k)$, and the gradients w.r.t. A and B. An one dimensional minimization was carried out along the negative gradient direction by a recursive interval halving technique, and the optimum step size determined. The starting values are now updated and the procedure is repeated. Stopping criterion is less than 2% improvement in the objective function or sum of the absolute values

of the elements of the gradient w.r.t. A and B be less than 0.30 and 0.50 respectively.

The program was run with close starting values for A. Since B is unknown, it is set equal to A initially. The selection functions at the end of 28 iterations are shown in table 6.1. It is seen that the convergence as far as the selection functions are concerned is quite slow. This manifests itself very adversely (in that it takes an extravagant amount of computation)when the starting guesses are off by about 20% or more. It appears that the lack of knowledge of the initial settings of B is mainly responsible for this situation.

One aspect of the method is worth pointing out. Since we are identifying the whole A matrix, by determining the corresponding continuous system matrix, we can evaluate all the breakage parameters without the assumption of normalizability. The entire "Z matrix" thus evaluated is shown in table 6.2 and indicates that the assumption about the normalizability of breakage parameter is justifiable.

Fig. 6.1 gives the minimisation of performance index J and the monotonic behaviour may be noted.

The covariance matrices P, R and F are next calculated using equations 6.2.10 and 6.2.11 and are shown in table 6.3, 6.4, 6.5.

6.4 Discussion

From the "most likely" estimates of the covariance matrices P, R and F, it is seen that the driving noise is much small compared with the measurement noise. Also, the variance of error in the estimated output and the measurement error covariance are nearly equal. This indicates that the model can be simplified considerably without serious error by ignoring the driving noise. It may be noted that this is in conformity with the earlier model in chapter III and IV which on the whole yield comparable results.

Due to the slow convergence of the algorithm, it has not been found possible to verify whether we will get consistent results for large errors in the initial guesses. The results at the end of 26 iterations for such a case is shown in table 6.1. All the same it is noted that the parameters approach the previous values.

CONVERGENCE OF SELECTION FUNCTIONS
MILL FEED 3300 GMS

Size Index	CASE I Start	At End of 28 Itrns	CASE II Start	At End of 26 Itrns
1	0.60	.6136	0.80	.6791
2	0.55	.5530	0.70	.6423
3	0.45	.4578	0.60	.5852
4	0.35	.3562	0.50	.4910
5	0.25	.2546	0.35	.3326
6	0.20	.2002	0.30	.3064
7	0.10	.1001	0.15	.1387
8	0.10	.1012	0.15	.1466
9	0.10	.0996	0.15	.1439
10	0.06	.0595	0.10	.1026
11	0.10	.0996	0.10	.1034
12	0.06	.0598	0.08	.0813
ERSUM x 10^4	139.94	88.792	4057.46	183.74

ERSUM = SUM OF SQUARES OF ERROR

TABLE 6.2

THE BREAKAGE PARAMETER MATRIX z_{ji}

j / i	1	2	3	4	5	6	7	8	9	10	11	12
2	.4428											
3	.2146	.4355										
4	.1010	.2113	.4340									
5	.0591	.0987	.2088	.4324								
6	.0454	.0613	.1023	.2132	.4372							
7	.0297	.0422	.0578	.0994	.2109	.4384						
8	.0257	.0285	.0429	.0580	.0988	.2111	.4343					
9	.0169	.0263	.0303	.0457	.0617	.1040	.2167	.4410				
10	.0134	.0175	.0273	.0317	.0472	.0635	.1064	.2170	.4443			
11	.0117	.0132	.0174	.0271	.0315	.0470	.0637	.1046	.2171	.4450		
12	.0108	.0117	.0132	.0175	.0273	.0317	.0479	.0629	.1079	.2180	.4431	

TABLE 6.3

ERROR COVARIANCE P OF ESTIMATED OUTPUT y

(MULTIPLIED BY WEIGHTING FACTOR 10^6)

Size Index	1	2	3	4	5	6	7	8	9	10	11	12
1	100.27	-0.887	-15.50	-6.026	-12.79	-0.638	-17.77	-4.929	-6.253	-2.323	-1.306	-2.531
2	-0.885	20.61	3.111	-1.182	-4.060	-4.827	-2.614	-4.413	-2.866	0.576	-0.787	-1.224
3	-15.50	3.111	3.901	-0.488	1.188	-0.901	3.062	0.112	0.434	0.451	0.051	0.200
4	-6.026	-1.182	-0.488	3.343	1.587	1.754	-1.148	1.127	0.137	-0.227	-0.012	-0.036
5	-12.79	-4.060	1.188	1.586	2.784	1.446	2.439	1.529	1.388	0.055	0.262	0.500
6	-0.638	-4.827	-0.901	1.754	1.446	3.659	-2.269	1.643	0.278	-0.270	0.155	0.168
7	-17.77	-2.614	3.062	-1.148	2.439	-2.269	7.677	0.610	2.221	0.553	0.406	0.847
8	-4.929	-4.413	-0.112	1.127	1.529	1.644	0.610	1.524	0.737	-0.112	0.235	0.339
9	-6.253	-2.866	0.434	0.133	1.388	0.278	2.221	0.733	1.175	0.189	0.249	0.461
10	-2.324	0.576	0.451	-0.227	0.057	-0.270	0.553	-0.112	0.189	0.285	0.088	0.129
11	-1.306	-0.788	0.051	-0.012	0.262	0.155	0.406	0.235	0.249	0.088	0.102	0.135
12	-2.532	-1.224	0.200	-0.036	0.500	0.167	0.847	0.339	0.461	0.129	0.135	0.215

TABLE 6.4

OUTPUT NOISE COVARIANCE MATRIX R

(MULTIPLIED BY WEIGHTING FACTOR 10^6)

Size Index	1	2	3	4	5	6	7	8	9	10	11	12
1	97.05	-0.833	-14.99	-5.865	-12.39	-0.655	- 17.16	-4.780	-6.050	-2.245	-1.260	-2.449
2	-0.766	20.58	3.090	-1.191	-4.070	-4.826	- 2.625	-4.410	-2.860	0.573	0.788	-1.286
3	-15.04	3.089	3.824	-0.508	1.134	-0.892	2.972	0.129	0.405	0.439	0.045	0.189
4	-5.159	-1.173	-0.623	3.296	1.474	1.751	- 1.314	1.082	0.075	-0.247	0.024	-0.059
5	-12.32	-4.062	1.113	1.566	2.725	1.450	2.345	1.508	1.357	0.045	0.256	0.488
6	-0.366	-4.802	-0.939	1.735	1.405	3.651	-2.320	1.624	0.257	-0.275	0.151	0.160
7	-18.07	-2.654	3.099	-1.121	2.487	-2.251	7.720	0.635	2.244	0.558	0.411	0.857
8	-4.600	-4.412	-0.164	-1.111	1.487	1.645	0.545	1.508	0.711	-0.120	0.231	0.330
9	-6.400	-2.874	0.456	0.145	1.419	0.283	2.240	0.743	1.186	0.192	0.251	0.465
10	-2.490	0.577	0.477	-0.218	0.078	-0.270	0.586	-0.104	0.200	0.289	0.091	0.133
11	-1.390	-0.787	0.065	-0.007	0.274	0.155	0.423	0.240	0.255	0.090	0.105	0.137
12	-2.630	-1.227	0.215	-0.029	0.515	0.170	0.866	0.345	0.468	0.131	0.136	0.217

TABLE 6.5

INPUT NOISE COVARIANCE MATRIX F

(MULTIPLIED BY WEIGHTING FACTOR 10^6)

Size Index	1	2	3	4	5	6	7	8	9	10	11	12
1	.4392	.0183	-.1402	.5981	.0215	.2650	-.8811	.1510	-.3445	-.2326	-.1261	-.1789
2	-.1464	.0048	.0116	.0371	.0198	.0292	-.0334	.0108	-.0091	-.0005	-.0017	.0042
3	-.0464	.0141	.0178	-.0901	-.0089	-.0399	.1195	-.0265	.0464	.0347	.0180	.0242
4	-.7538	-.0137	.1220	.0129	.0079	-.0113	.1880	.0326	.0671	.0269	.0159	.0288
5	-.1152	.0016	.0305	-.0795	.0048	.0418	.1381	-.0173	.0515	.0315	.0174	.0259
6	-.2966	-.0248	.0468	.0181	.0458	.0011	.0696	.0215	.0252	.0045	.0044	.0098
7	.8798	.0440	-.1166	-.1865	-.1410	-.0713	-.0006	.0900	.0033	.0275	.0121	.0096
8	.2019	.0004	.0383	-.0232	.0227	-.0193	.0891	.0014	.0315	.0152	.0086	.0145
9	.3615	.0108	-.0471	-.0677	-.0534	-.0260	-.0037	-.0322	.0000	.0081	.0037	.0028
10	.2486	-.0017	-.0379	-.0297	-.0337	-.0055	-.0271	-.0163	-.0077	.0003	.0000	-.0020
11	.1338	.0004	-.0196	-.0171	-.0184	-.0048	-.0119	-.0092	-.0035	.0002	.0001	-.0008
12	.1849	.0038	-.0258	-.0298	-.0270	-.0104	-.0095	-.0150	-.0026	.0023	.0010	.0001

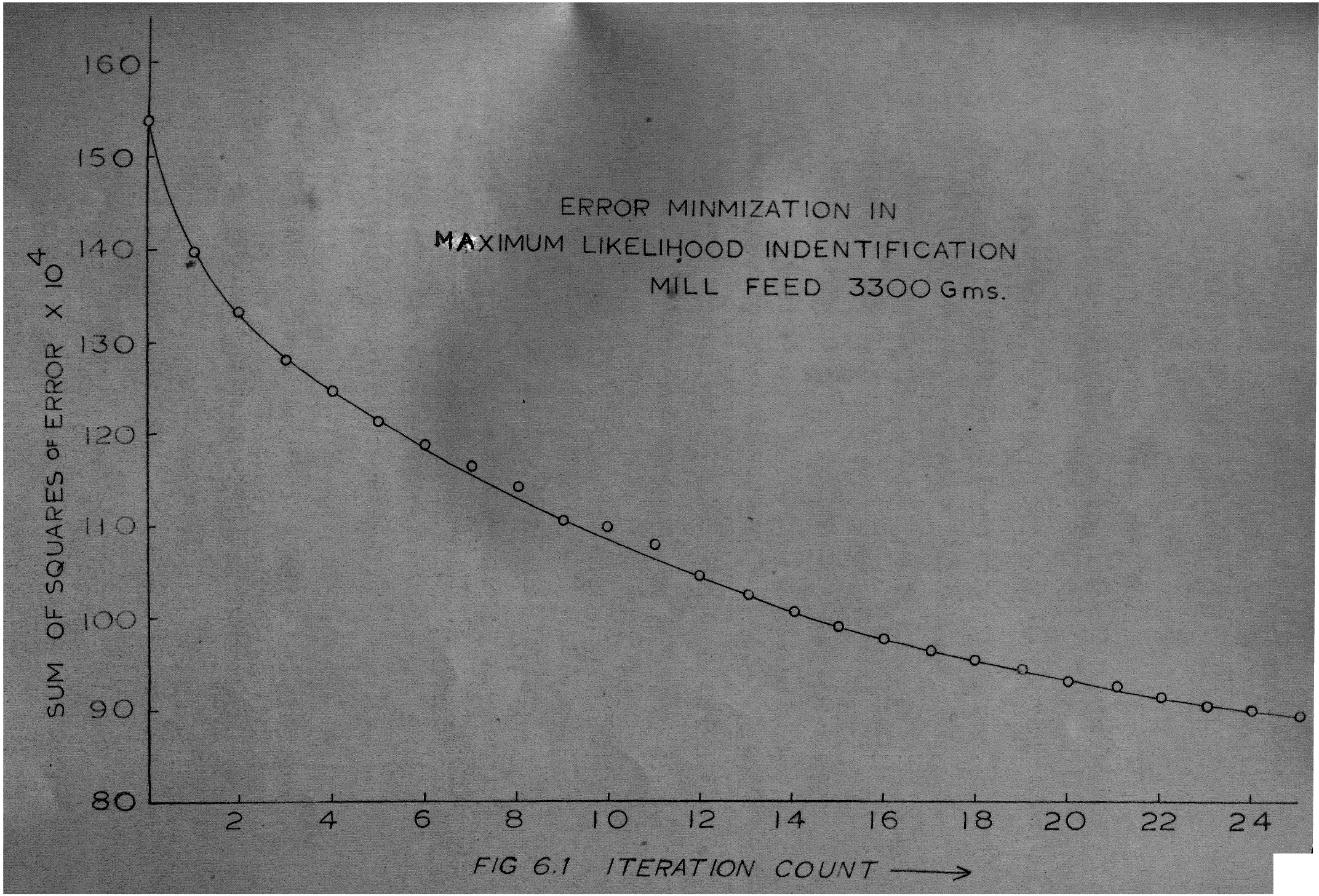

ERROR MINMIZATION IN
MAXIMUM LIKELIHOOD INDENTIFICATION
MILL FEED 3300 Gms.
SUM OF SQUARES OF ERROR X 10^4
160
150
140
130
120
110
100
90
80
2
4
6
8
10
12
14
16
18
20
22
24
ITERATION COUNT

FIG 6.1

CHAPTER VII

STOCHASTIC APPROXIMATION

7.1 Introduction

An interesting approach to the problem of parameter identification characterised by a large amount of 'dirty' data and the absence of any knowledge of statistical information is through the methods of Stochastic Approximation. The Stochastic Approximation algorithms often referred to as learning procedures may be considered as the stochastic analogs of their deterministic counterparts. An identification problem in the framework of stochastic approximation may be stated as follows:

Let θ be the parameter we are trying to identify and $J(\theta)$ the function to be minimised, which however is unknown. But we have experimental results where we observe $y(\theta)$ and let $g(\theta)$ be a known function of $y(\theta)$ and θ such that $E\left[g(\theta)/\theta\right] = J(\theta)$. So essentially we look for an algorithm by which we can determine $\theta = \theta^{\circ}$ which minimises $J(\theta)$ by requiring only measurements of the random variable $y(\theta)$.

7.2 The Kiefer Wolfowitz Algorithm

As seen in the last section we are interested in finding out the extremum of a function $J(\theta)=E\left[g(\theta)/\theta\right]$

where we are able to determine $g(\theta)$ given any nominal value of the parameter θ. But no knowledge of the statistics of the random variable $g(\theta)$ is available. The Kiefer-Wolfowitz scheme uses the estimate of the gradient $\hat{\nabla}_\theta [g(\theta)/\theta]$ computed from neghbouring values of $g(\theta)$ by finite perturbation of θ. When θ is a vector (say of dimension n) one component is perturbed at a time. We observe the random samples of $g(\theta)$ which are $g(\theta_k \pm c_k e_j)$, $j = 1,2 \ldots\ldots$ n where c_k is a suitably chosen positive sequence, and $e_1, e_2 \ldots\ldots e_n$ are the n dimensional unit

Vectors :

$$e_1 = (1, 0, \ldots\ldots 0)^T$$

$$e_2 = (0, 1, \ldots\ldots 0)^T$$

$$\vdots$$

$$e_n = (0, 0, \ldots\ldots 1)^T$$

The subscript k indicates a sequence for the recursive updating of θ as indicated below:

$$\text{Thus} \quad \hat{\nabla}_\theta\, g(\theta_k) = \sum_{j=1}^{n} e_j \frac{1}{2c_k} \left[g(\theta_k + c_k e_j) - g(\theta_k - c_k e_j) \right] \qquad (7.2.1)$$

Using the estimate of the gradient, the computing scheme can be written as

$$\theta_{k+1} = \theta_k - d_k \hat{\nabla}_\theta \left[g(\theta_k) \right] \qquad (7.2.2)$$

where d_k is a suitably chosen sequence.

For convergence, the sequences c_k and d_k have to satisfy the following conditions:

$$\sum d_k = \infty \quad ; \quad \sum d_k c_k < \infty$$
$$\lim_{k \to \infty} c_k = 0 \quad ; \quad d_k < d_1 c_k^2 , \; 0 < d_1 < \infty$$
$$\text{and} \quad (d_k / c_k)^2 < \infty \qquad (7.2.3)$$

The conditions that should be satisfied for convergence of KW scheme as given by Venter (1967) are:

1. $J(\theta)$ and its second order derivatives are bounded in R^n.

2. θ° is a local minimum of $J(\theta)$. ie, for some $\varepsilon > 0$, we have

$J(\theta^\circ) \leq J(\theta)$ for all $\theta = \theta^\circ$ in the set $\{\theta : \| \theta - \theta^\circ \| < \varepsilon\}$

3. For every $\theta \in R^n$, $\hat{\nabla}_\theta J(\theta) \neq 0$ if $\theta \neq \theta^\circ$.

That is, θ° is the only stationary point of $J(\theta)$.

4. When $J(\theta)$ is redifined as $J(\theta) = g(\theta) + v(\theta)$ with $E\{v(\theta)\} = 0$, $v(\theta)$ should satisfy the condition

$$E\{v^2(\theta)\} < \infty \quad \forall \quad \theta \in R^n. \tag{7.2.4}$$

Venter has shown that if these conditions are met, the algorithm given by (7.2.2) with gains satisfying (7.2.3) converges with probability one to either θ° or infinity.

7.3 Computational Considerations

In applying the algorithm to our problem, $g(\theta)$ is chosen as the sum of squares of error between the nominal trajectory given by the parameter value θ and the experimental data. This enables comparison with the results of the previous chapters.

The sequences c_k and d_k are chosen so as to satisfy (7.2.3). The usual choice is $d_k = d_1 / k$ and $c_k = c_1/k^{\alpha}$ where d_1 and c_1 are positive and $0 < \alpha < 0.5$. We choose $\alpha = 0.4$, and c_1 adjusted so that the perturbance on the performance index is neither too large nor too small. d_1 was chosen by trial and error to get rapid convergence, and range from 0.5 to 40.0 for the 12 parameters. It is seen that apart from the rate of convergence, the choice of

d_1 within reasonable limits does not affect the eventual result seriously, provided there is no violent oscillation for any parameter values.

It is noted that although the algorithm is quite simple, it is computationally inefficient. For each perturbation of every parameter, the nominal trajectory has to be evaluated and change in performance index computed. Hence it was decided to identify only the selection functions by this method, keeping the breakage parameters at the value given by the Quasilinearization method.

7.4 Results

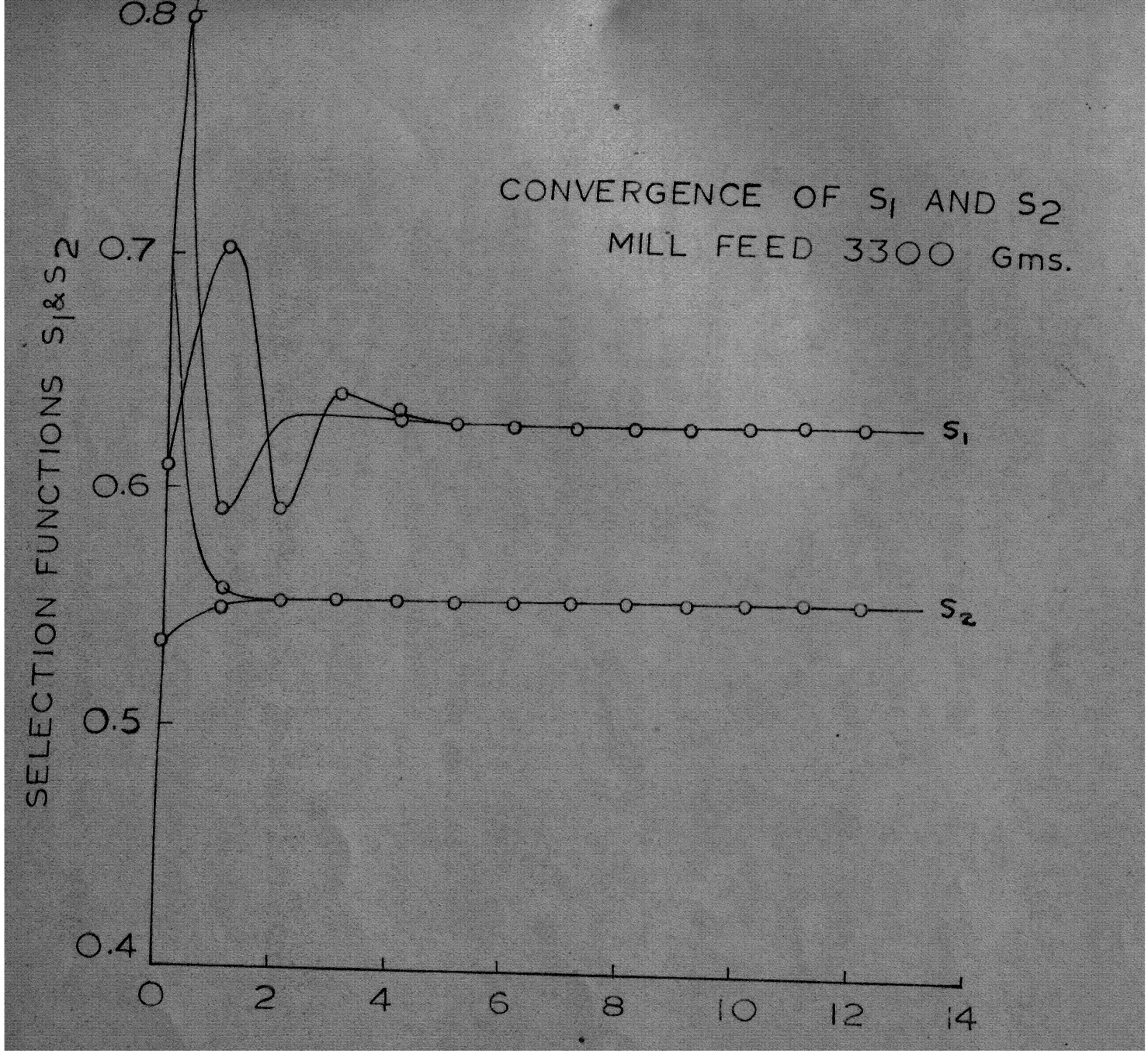
Table 7.1 summarizes the numerical results of the estimation of the selection functions for two different starting values. It is seen that by adjusting the coefficients d_k, it is possible to obtain faster convergence even if the initial settings are far off from the optimum values. The minimization of the performance index is also given. It can be observed that the results for the two different starting guesses are more or less consistant. The convergence of S_1 and S_2 are shown in figure 7.1.

TABLE 7.1

IDENTIFICATION OF SELECTION FUNCTIONS

MILL FEED - 3300 GMS

Size Index	Z	Case I		Case II	
		Start	At End of 24 Itrns	Start	At End of 27 Itrns
1	.4386	.6095	.6314	0.80	.6318
2	.2137	.5453	.5531	0.70	.5542
3	.1008	.4675	.4583	0.60	.4607
4	.0607	.3828	.3582	0.50	.3598
5	.0442	.2983	.2643	0.35	.2171
6	.0310	.2189	.1995	0.30	.2007
7	.0248	.1504	.1122	0.15	.1137
8	.0173	.0973	.1142	0.15	.1145
9	.0131	.0633	.1042	0.15	.1035
10	.0112	.0539	.0621	0.10	.0635
11	.0102	.0741	.0961	0.10	.0954
12	--	.1294	.0507	0.08	.0497
Sum of Squares of Error x 10^4		27.53	12.37	489.70	12.42


CONVERGENCE OF S_1 AND S_2
MILL FEED 3300 Gms.
SELECTION FUNCTIONS S_1 & S_2
0.8
0.7
0.6
0.5
0.4
0
2
4
6
8
10
12
14
S_1
S_2

CHAPTER VIII

CONCLUSION

The problem of identification of a dynamical system in noisy environment is essentially that of stochastic observability of a set of unknown parameters, given a set of related measurements. Although several procedures have been suggested in the literature for the identification of systems, in the absence of any explicit creiteria for the observability of the parameter, the feasibility of these approaches must be judged in the light of their computational complexity, the amount of numerical experimentation required and the trend towards convergence of the estimates for a given data. Clearly any conclusion regarding these factors has to be restricted to a specific problem. It is in this premise that this thesis is investigated where an attempt has been made to study the question of identification of a batch grinding mill - a problem of considerable engineering importance.

Different approaches for the identification of the given system were discussed in the preceeding chapters. A comparitive evaluation of the procedures is considered appropriate. At the same time the question as to whether comparisons will be meaningful has also to be examined, since every method has one or other distinctive aspect associated with it.

As far as the performance index is concerned, there is close similarity between the methods of Recursive Lease Squares, Quasilinearization and Stochastic Approximation. So it is natural to expect comparable results from the three schemes. As far as the selection function goes, it is seen that this expectation is satisfied to a large extent. It is also worth pointing out that in all the three methods, the initial guesses for the parameters is not a serious issue. Except that more computer time will be involved, a large error in the initial setting does not materially alter the final results. A still more important observation relevant here is that the available data which is limited and of non-uniform distribution does not pose any major difficulty either in the implementation or convergence of the program. It was not considered necessary to resort to any formof interpolation. As far as the specific problem is concerned, it is seen that Quasilinearization promises to be most suitable in many respects. Identification of both selection and breakage functions is achieved without lot of involved computation. Convergence is monotonic and rapid.

The method of Invariant Imbedding stands apart in many respects. Briefly the identification scheme calls

for more information about the process than is available. It is possible that numerical experimentation can answer some of the specific issues like choosing weighting matrices etc satisfactorily in a low dimensional problem, but not necessarily so otherwise. It is in fact possible to pose the question of best choice of weighting matrices in a situation where noise statistics are nor available as a problem in itself, and has not been answered here. On the whole it is felt that the results of this chapter only marginally complement rather than supplement what is already known.

The maximum likelihood approach differs from the rest in that we seek to identify the discrete plant matrix in full, and estimate the associated noise convariances. A more comlete description is thus obtained. The performance index is apparently the same as before, but is evaluated in a different way. The estimate of the covariance of the driving and measurement noise is considered significant, since it enables comparison of the relative strengths. As has been mentioned earlier, the assumption of Gaussian distribution is inappropriate in our problem. It will be interesting to see whether some non Gaussian density can be employed, but it is likely to introduce mathematical complications.

Table 8.1 summarises the important results of the different schemes for easy comparison.

Briefly the main results of this thesis are the following:

1. The linear time invariant model that was chosen appears to be quite adequate to describe the system behaviour in the time interval of interest.

2. Since no general results concerning convergence or uniqueness of parameter values are obtained, the fact that we are able to obtain reasonably close estimates using different procedures is significant. Heuristically speaking this solves the problem more or less satisfactorily.

It might be possible to increase the accuracy of the computations by using more sophisticated techniques like higher order predictor-corrector schemes for solving the differential equations, improved methods for matrix inversion etc. Also approaches like instrumental variables or time series analysis can considered. But basically it seems there can be no escape from the same sort of difficulties as above, posed by moderately large number of parameters to be identified with limited data coupled with lack of pertinent statistical information.

OF DIFFERENT METHODS

Size Index	Recursive Least squares	Quasilineari-zation	Invariant Imbedding	Macimum Likelihood	Stochastic Approximation
1	.6203	.6214	.6179	.6136	.6314
2	.5457	.5526	.5483	.5530	.5531
3	.4416	.4525	.4457	.4578	.4583
4	.3462	.3519	.3436	.3562	.3502
5	.2518	.2585	.2460	.2542	.2643
6	.1908	.1941	.1957	.2002	.1995
7	.1046	.1117	.1136	.1001	.1122
8	.1008	.1140	.1134	.1012	.1142
9	.0950	.1087	.1135	.0996	.1042
10	.0482	.0629	.0689	.0595	.0621
11	.0761	.0973	.0791	.0996	.0961
12	.0273	.0502	.0539	.0598	.0507
Per. Index (weighted in col. 4)	18.5×10^{-4}	14.1×10^{-4}	0.1535	88.79×10^{-4}	12.38×10^{-4}
Time taken in FORTRAN	5 MIN 12 SEC.	7MIN 45SEC	6 MIN 40 SEC	27 MIN 14 SEC	15 MIN 52 SEC

REFERENCES

1. Berlioz L.M., Kinetics of Dry Grinding, M.S. Thesis University of California, Berkeley, 1966.
2. Charles R.J., Energy Size Reduction Relationships in Comminution, AIME Trans. 1957 pp 80.
3. Detchmendy D.M., Identification, Estimation and Control of Nonlinear Systems, Ph.D. Thesis Purdue University 1965.
4. Herbst J.A. and Fuerstenau D.W., An Interpretation of the Zero Order Production of Sizes in Comminution Processes and its Implication in Simulation, AIME Trans. 1968 pp 538.
5. Kapur P.C., Kinetics of Batch Grainding. Part A : Reduction of the Grainding Equation, AIME Trans. Vol. 250, 1971.
6. Kashyap R.L. Maximum Likelihood Identification of Stochastic Linear Systems, IEEE AC, Vol. AC-15 No.1 Feb; 1970, pp 25-34.
7. Kashyap Blaysden and Fu, "Stochastic Approximation" in "Adaption, Learning and Pattern Recoginition" Eds. J.M. Mendel and K.S. Fu, Academic Press 1970.
8. Lee E.S., Quasilinearization and Invariant Imbedding, Academic Press, 1968.

9 Loginov N.V., Methods of Stochastic Approximating, Survey Paper in Automation and Remote Control 1966, pp 706-728.

10 Nahi N.E., Estimation Theory and Application, John Wiley, 1969.

11 Rao P.R.K., Identification and Estimation of Linear Non-stationary Dynamic Systems. Ph.D. Thesis., Dept. of Elec. Engineering, IIT Kanpur, 1970.

12. Reid, K.J., A Solution of the Batch Grinding Equation, Chemical Engg. Sci., 1965, pp 953-963.

13. Sage A.P., Optimum Systems Control, Prentice Hall 1968.

14. Sakrison D.J., "Stochastic Approximation" in Advances in Communication Systems. Ed. A.V. Balakrishnan, Vol.II, 1966.

15. Schuhmann R. Jr., Principles of Comminution-Size Distribution and Surface Calculations AIME TP 1189, Mining Technology July 1940.

16. Venter J.H., On the Convergence of the KW Approximation procedure, Annals of Mathematical Statistics, No.4, pp 1031, 1967 b.